RESTRICTED VERB PHRASE COLLOCATIONS IN STANDARD AND LEARNER MALAYSIAN ENGLISH

A thesis submitted in partial fulfilment of the requirements for

the Degree

of Doctoral of Philosophy of Linguistics

in the University of Canterbury

by Hasliza Abdul Halim

University of Canterbury

2014

Supervisory team:

Emeritus Professor Dr Koenraad Kuiper

Professor Dr Thomas Klee

Abstract

The English used in Malaysia is one of the varieties of New Englishes and this variety has emerged due to the spread of English around the world (Platt, et al., 1983; Pillai, 2006). In the case of Malaysia, Malay is the national language and standard English exists to be the language of an elite (Bao, 2006), also as a language of interaction. Over years of playing its various roles as a language of interaction, there has emerged a variety of English that is distinctively Malaysian (Asmah, 1992). Baskaran (2002) points out that English is now adopted and adapted in the linguistic ecology of Malaysia, and all Malaysians should be proud of it with all its local 'nuances and innuendos'. Malaysian English today is 'a rich tapestry of a typical transplanted variety of English'. Malaysian English (ME) is one of the new varieties of English, with some distinct features include the localized vocabulary, pronunciation, grammar as well as pragmatic features (Pillai, 2006; Pillai and Fauziah, 2006, p.39). The present study has embarked on a specialised study of vocabulary. In particular, it examined the English collocations produced by non-native speaker English users in Malaysia. The study provided insight into the nature of the internal norms of English used in Malaysia to see how these English restricted collocations being used by this group of learners. The investigation focused on the learners' productive knowledge of Verb-Noun collocations of their written English with the impact of exposure and frequency. Nesselhauf (2003) has the opinion that verb-noun combinations are the most frequently mistaken so they should perceive particular attention of learners.

Investigating collocation in English language learning is paramount as such study may inform us on the use of restricted collocations in English language teaching and learning in Malaysian context. The findings in Chapter 4 and 5 suggest that the frequency of the cloze verb does have an effect as predicted by Kuiper, Columbus and Schmitt (2009). This is so because frequency is a measure of likely exposure. The more frequent an item is in corpora, the more likely a learner is to be exposed to it. What is needed is a much more nuanced notion of exposure. The findings in Chapter 6 proves that the malformed collocations make sense could be a way of making the World English perspective relevant after all. A new testing approach is proposed; semantic plausibility metric, which is used as a tool for this study, can be useful used as a measure of vocabulary acquisition as well as looking at learners' test taking strategies. The findings of the present research on Malaysian English collocations contribute new knowledge to the existing understanding and literature on the acquisition of collocations.

Keywords: Collocations, restricted collocations, exposure, frequency, malformed collocations, semantic plausibility metric

Acknowledgements

My utmost gratitude to Allah for answering my prayers and giving me strength to complete this dissertation.

It is a pleasure to thank many people who made this thesis possible.

First and foremost I offer my sincere gratitude to my first supervisor, Emeritus Professor Koenraad Kuiper, who has supported me throughout my thesis with his patience and knowledge. I would like to attribute the level of my doctoral degree to his encouragement, enthusiasm, inspiration and countless efforts, and without him this thesis would not have been completed or written.

I would like to thank Professor Thomas Klee who provided sound advice and helped with various statistical analyses. Throughout my thesis-writing period, he provided encouragement, good teaching and lots of good ideas.

In the various language laboratories and handling computer software, I have been aided in running the software by Robert Fromont, a software programmer who helped with the Wordsmith Tools 5, CELEX and all other related work in obtaining the verb frequency list. The smooth running of the lab work and data collection activities were made possible by the support from Professor Jen Hay, Professor Beth Hume, Professor Lucy Johnston, Dr Heidi Quinn, Dr Kevin Watson, Professor Susan Foster-Cohen, Jacqui Nokes, Dr Patrick la Shell, Dr Lynn Clark, Emma Parnell, Dr Irina Loginova, Victoria Escaip, Dr Vica Papp, Darcy Rose, Bruce Peyno (ICT), Susan Boutery, Anandi Eichenberger and everybody at Linguistics Department, the

New Zealand Institute of Brain and Behaviour (NZILBB), and SAPS. In many ways I have learnt much from Matt Ward, who has fascinated me with his ability in mathematics and statistics.

I am indebted to many people for providing guidance related to corpus development. I am especially grateful to Dr Georgie Columbus who was very helpful providing advice and suggestions on corpus tagging, lemmatization and getting verb frequency list. Professor Paul Nation also deserves special thanks for the suggestions he made in reference to corpus building, as well as Dr Mike Scott who has helped me through emails running Wordsmith Tools 5.0. I wish to thank Dr Paul Rayson for his advice and communications through emails regarding corpus tagging work. A special thank goes to Dr Alison Kuiper whose support and insight were useful throughout the process. I am very grateful to Professor Dr. Annelies Hacki-Buhofer for her feedbacks at the EUROPHRAS Maribor 2012.

The research was supported by University of Canterbury grants from the School of Languages, Cultures and Linguistics and the School of Social and Political Sciences.

This study has become possible through the 6-years of news reports of News Straits Times, NST archive provided by NST press. The permission to get free access of 6-years of news reports was approved by Dato' Syed Nadzri Syed Harun, the NST Group Editor (see Appendix A). The corpus of the present study is named after the provider of the texts, i.e. NST corpus. Thanks also go (a million or so) to my employer, Universiti Utara Malaysia, an academic institution in Malaysia which has funded the study for these few years. The full study scholarship and allowances have

made the journey very smooth. And I am indebted to the staff as well, too many to mentioned here but among are Professor Dato' Wira Dr. Mohamed Mustafa Ishak, Professor Rosnah Awang Hashim, Professor Abdul Razak Salleh, Professor Ku Rahana, Dr Siti Jamilah, Associate Professor Nuraini Yusoff, Dr Aspalila and all UUM staff from Unit Pengajian Lanjutan and Language Centre.

I was lucky to be awarded a study grant funded by Malay Chair from the Victoria University of Wellington (see Appendix B) and the special thanks is dedicated to Professor Wan Rafaei, Cheryl McGettigan and Deddy Pasareken. I would like to thank the Human Ethics Committee of the University of Canterbury, and Economic Planning Unit of Prime Minister's Department, Malaysia for the support and approval for the data collection purposes (see Appendix C).

I am also indebted to many colleagues from Universiti Utara Malaysia for providing me support throughout the data collection period. Special thanks to Baizura and Eny who have been there for me through thick and thin. I am also grateful to all the Malaysian community in Christchurch especially students of University of Canterbury and other Malaysian students in New Zealand who were involved in the study. Also thanks to all school principals and students involved in the study; those are from Setiabudi High School, International Islamic School Gombak, Sultanah Asma High School and Sultan Abdul Halim High School (Jenan), Malaysia.

I must confess that I am truly indebted to Professor Che Su Mustafa, Dr Hassan Abu Bakar and Professor Ku Ruhana who believe that 'an ordinary teacher can dream big

and go beyond the ordinary'. They have made the journey possible to begin, and believe that it would end with a good ending.

Finally, I thank my husband, Suhaimi, and my two little heroes; Haziq and Haqiem, who understand me, and help me stay sane through difficult years. Both my parents, Hj Abd Halim and Hjh Napishah, thank you for your support and prayers throughout my study. To my siblings, Haslinawati, Hasrul and Haswad and the entire extended family, thank you for providing a loving environment for me. My journey wouldn't have started or even ended without their love and support. To them I dedicate this thesis.

Allow me to dedicate a single paragraph to a place that I called second home; the quake city of Christchurch, where the writing of this thesis took place. I would love to share a sketch of how my PhD journey was very much affected by the quakes.

The earthquakes in September 2010 and February 2011 and series of thousands aftershocks have indeed affected not just my life, but the entire family. Everyone is hurting in some way over the trauma that we have been through. Whether we jump at the sign of quakes, refuse to go into buildings more than a single story high, or wear a hat or helmet in shopping malls - we all have scars left behind. Regardless of emotional scars, the spirits of the Christchurch City and the highly spirited Cantabrians have inspired us to move on. How we missed enjoying lunch at the Square while looking at visitors taking photos of the (older) Cathedral and enjoying the street life. I remember we snapped several photos along the grassy banks of the

Avon. But to us, even though the experiences of quakes were totally horrific, Christchurch will stay close to the heart and mapped as the ‘green zone’ and, never the ‘red zone’.

Hasliza Abd Halim

February 2014

Canterbury

Papers presented at these conferences:

1. A poster presentation at First National Postgraduate Research Colloquium 2011(1NPRC)-Research in Malaysia, Otago University, New Zealand on July 4, 2011. Organizer- Chair of Malay Studies, Victoria University of Wellington.

2. A poster presentation at Linguistics Society of New Zealand (LingSoc) conference at Victoria University of Wellington on November 21, 2011. Organizer- Linguistics Society of New Zealand (LingSoc).

3. An oral presentation at EUROPHRAS Maribor 2012 at University of Maribor, Slovenia on 27-31 August 2012. Organizer- European Society of Phraseology (EUROPHRAS).

4. An oral presentation at School of Languages, Cultures and Linguistics Postgraduate Research Seminar at School of Languages, Cultures and Linguistics, University of Canterbury on November 12, 2012. Organizer- School of Languages, Cultures and Linguistics, University of Canterbury.

Award:

Awarded a study grant funded by Malay Chair from the Victoria University of Wellington.

Member:

A member of Linguistics Societyof New Zealand (Inc) www.nzlingsoc.org since November 2011.

A member of EUROPHRAS since August 2012.

Table of Contents

Abstract ii
Acknowledgement iii
Table of Contents ix
List of Tables xiii
List of Figures xvii
List of Appendices xviii

CHAPTER ONE INTRODUCTION 1
1.1 Background 1
1.2 The importance of the study 4
1.3 Aims of the study 8
1.4 The shape of the thesis 16

CHAPTER TWO REVIEW OF THE RELEVANT LITERATURE 18
2.1 Introduction 18
2.2 Phraseology 18
2.3 What is collocation? 24
2.3.1 The importance of collocations in language learning 26
2.4 Restricted collocations 31
2.5 The acquisition of phrasal vocabulary by native speakers and learners 34
2.6 The importance of frequency approach in lexical studies 37
2.7 Vocabulary acquisition and the effects of exposure 43
2.8 Vocabulary acquisition in Malaysia 46
2.8.1 Language policy and situation in Malaysia 49
2.9 Theoretical framework of the study 53
2.10 Chapter summary 58

CHAPTER THREE RESEARCH METHODOLOGY 60
3.1 Introduction 60
3.2 Background:Creating the NST Corpus 64

3.2.1 Permissions……65
3.2.2 Spoken and written corpus……65
3.2.3 Corpus size……66
3.2.4 Balanced corpus……67
3.3 Tagging and lemmatization……..69
3.3.1 Tagging……..69
3.3.2 Lemmatization……...73
3.4 Creating a wordlist from NST corpus……..……74
3.4.1 How the verb frequency count was established……80
3.4.2 Summary of the corpus building……...85
3.5 Constructing a cloze test……..87
3.5.1 Introduction……...……87
3.5.2 Objectives…….....90
3.5.3 Procedure…….....90
3.5.4 Categorization of verb frequency…….....91
3.6 Checking the frequency effects of both verb frequency lists for Study 1 and Study 2……97
3.6.1 Evaluating the studies…….....102
3.7 Testing the reliability of both cloze testing sets for Study 1 and Study 2…..…104
3.8 Chapter summary.……...105

CHAPTER FOUR-STUDY 1: THE ACQUISITION OF MALAYSIAN ENGLISH RESTRICTED COLLOCATIONS.....107

4.1 Introduction.....107
4.1.1 Research questions and hypotheses…..……..108
4.2 Methodology ……..109
4.2.1 Measures……..109
4.2.2 Participants……..110
4.3 Procedure..……..115
4.4 Results ……...……117
4.4.1 Participant groups and head verb frequency…..…….……118
4.4.2 The role of frequency of restricted collocations on acquisition……..123

CHAPTER FIVE -STUDY 2: THE ACQUISITION OF RESTRICTED COLLOCATIONS OF NEW ZEALAND ENGLISH…… 126
5.1 Introduction 126
5.1.1 Research questions and hypotheses..……..127
5.2 Methodology ……..129
5.2.1 Measures……..129
5.2.2 Participants……..130
5.3 Procedure..……..131
5.4 Results……134
5.4.1 Year of study and head verb frequency……...134
5.4.2 Correlation results of year of study, age and month of learning…...137
5.4.3 Multiple regression results of year of study, age and month of learning……140
5.4.4 The role of frequency of restricted collocations on its acquisition….144
CHAPTER SIX THE ANALYSIS OF THE SELECTION OF NON-IDIOMATIC HEADS 146
6.1 Introduction……...….146
6.2 Background…….……..…..…149
6.2.1The relationship between the exposure to a lexical item and the learning of that item……149
6.2.2 Native and non-native collocational knowledge…….151
6.3 Methodology…….……..152
6.3.1 Participants……..……152
6.3.2 Procedures……...154
6.4 Analysis……...…...169
6.5 Results and discussion……189
6.5.1 Individuals who have more idiomatic responses will have higher plausibility scores for other responses……..189
6.5.2 Case studies: samples of individuals set of responses…….192
6.5.3 Discussion……197
6.5.4 Conclusion……...202

CHAPTER SEVEN CONCLUSION, SUGGESTIONS AND TEACHING IMPLICATIONS 204

7.1 Introduction 204

7.2 Corpus approach..... 204

7.3 Exposure to the target language collocations..... 205

7.4 Frequency effects..... 210

7.5 Individual differences..... 212

7.6 Suggestions and teaching implications 213

7.7 Strength and limitations of the present study..... 217

REFERENCES 220

List of Tables

Table 2.1 Categorization of verbs in Kuiper, Columbus & Schmitt 40
Table 2.2 The subtypes of grammatical collocations.....56
Table 2.3 The subtypes of lexical collocations 57
Table 3.1 The frequency of RCs in BNC and NST corpora 63
Table 3.2 The tagset used for tagging the NST corpus.....70
Table 3.3 Statistics of NST corpus.....79
Table 3.4 Kuiper, Columbus & Schmitt's verb frequency bands.....92
Table 3.5 Study 1 test items ordered by head verbs' bands.....92
Table 3.6 Head verb frequency data and band allocation obtained through Wordsmith Tools 5.0.....94
Table 3.7 Head verb frequency data and band allocation obtained through CELEX .... lemmatization.....95
Table 3.8 Categorization of verb in NST corpus 96
Table 3.9 The frequency of head verbs and frequency of restricted collocations in NST corpus 101
Table 3.10 The frequency of head verbs and frequency of restricted collocationss in BNC 101
Table 3.11 The relationship between the frequency of the head verbs and the number of PLIs with that head in SAID.....103
Table 3.12 The reliability results for both test items 104
Table 3.13 Individual reliability results for each item of verb frequency.....104
Table 4.1 Number of respondents involved in the study 112
Table 4.2 Questionnaires obtained from native speakers of NZ English 114
Table 4.3 Questionnaires obtained from school students in Malaysia.....114
Table 4.4 Questionnaires obtained from university students in Malaysia 114
Table 4.5 Questionnaires obtained from Malaysian students studying in NZ 115
Table 4.6 List of 20 headverbs for the first study 116
Table 4.7 List of 20 RCs tested on students.....117
Table 4.8 Mean and standard deviation 119
Table 4.9 Summary of correlation results.....124

Table 5.1 Frequency data and band allocation of head verbs 132
Table 5.2 List of RCs used for Study 2 133
Table 5.3 Mean and standard deviation for scores in the cloze test by 3 groups of participants and 4-type head verb frequency 136
Table 5.4 The relationship between months of exposure and verb types 138
Table 5.5 Correlation between either year of study and age with other variables especially total and verb types 139
Table 5.6 Mean and standard deviation of year of study and age with total scores and verb types 140
Table 5.7 The mean score and standard deviations of DV; total score, and a set of IV; age, year of study and months of learning 143
Table 5.8 Predictors of total scores 143
Table 5.9 Summary of correlation results 144
Table 6.1 List of pharasal lexical items used in the test.....155
Table 6.2 List of responses for Case 1 156
Table 6.3 List of responses for Case 2 157
Table 6.4 List of responses for Case 3 157
Table 6.5 List of responses for Case 4 158
Table 6.6 List of responses for Case 5 159
Table 6.7 List of responses for Case 6 159
Table 6.8 List of responses for Case 7 160
Table 6.9 List of responses for Case 8 161
Table 6.10 List of responses for Case 9 162
Table 6.11 List of responses for Case 10 162
Table 6.12 List of responses for Case 11 163
Table 6.13 List of responses for Case 12 163
Table 6.14 List of responses for Case 13 164
Table 6.15 List of responses for Case 14 164
Table 6.16 List of responses for Case 15 165
Table 6.17 List of responses for Case 16 166
Table 6.18 List of responses for Case 17 166
Table 6.19 List of responses for Case 18 167

Table 6.20 List of responses for Case 19 168
Table 6.21 List of responses for Case 20 168
Table 6.22 Student 1(T1) responses 169
Table 6.23 Student 2(T2) responses 170
Table 6.24 Student 3(T3) responses 171
Table 6.25 Student 4(T4) responses 172
Table 6.26 Student 5(T5) responses 173
Table 6.27 Student 6(T6) responses 174
Table 6.28 Student 7(T7) responses 175
Table 6.29 Student 8(T8) responses 176
Table 6.30 Student 9(T9) responses 177
Table 6.31 Student 10(T10) responses 178
Table 6.32 Student 11(T11) responses 179
Table 6.33 Student 12(T12) responses 180
Table 6.34 Student 13(T13) responses 181
Table 6.35 Student 14(T14) responses 182
Table 6.36 Student 15(T15) responses 183
Table 6.37 Student 16(T16) responses 184
Table 6.38 Student 17(T17) responses 185
Table 6.39 Student 18(T18) responses 186
Table 6.40 Student 19(T19) responses 187
Table 6.41 Student 20(T20) responses 188
Table 6.42 Percentage of idiomatic responses, semantic plausibility mean and standard deviation of individuals taking the test 189
Table 6.43 Means and standard deviation for responses of 20 respondents 191
Table 6.44 Summary of correlation results for idiomatic responses and mean non-idiomatic responses for 20 respondents 191
Table 6.45 Student 5 (T5)- Anna's responses 192
Table 6.46 Band, sort order and frequency of the plausibility verbs responses 194
Table 6.47 Band 1 of the first 10 verbs extracted from Kilgarriff's lemmatized BNC frequency list 195

Table 6.48 Scores of three respondents representing the highest, middle and low group 196
Table 6.49 Eusoff's most semantically plausible responses.....197
Table 6.50 Reyna's most semantically plausible responses 197
Table 7.1 The highly plausible rated verbs 216

List of Figures

Figure 2.1 A common conceptualization of vocabulary frequency 42
Figure 2.2 Schmitt and Schmitt (2012) approach on vocabulary frequency.....42
Figure 2.3 Howarth's model of continuum.....54
Figure 3.1 An example of news report from NST in Plain text form 72
Figure 3.2 An example of a tagged text 73
Figure 3.3 Step 1 of NST corpus wordlist procedure 75
Figure 3.4 Step 2 of NST corpus wordlist procedure 76
Figure 3.5 Step 3 of NST corpus wordlist procedure 76
Figure 3.6 Step 4 of NST corpus wordlist procedure 77
Figure 3.7 Step 5 of NST corpus wordlist procedure 77
Figure 3.8 Step 6 of NST corpus wordlist procedure 78
Figure 3.9 Step 7 of NST corpus wordlist procedure 79
Figure 3.10 Step 1 of the RC frequency search in NST corpus 81
Figure 3.11 Step 2 of the RC frequency search in NST corpus 82
Figure 3.12 Step 3 of the RC frequency search in NST corpus 82
Figure 3.13 Step 4 of the RC frequency search in NST corpus 83
Figure 3.14 Step 5 of the RC frequency search in NST corpus 84
Figure 3.15 Step 6 of the RC frequency search in NST corpus 84
Figure 3.16 Step 7 of the RC frequency search in NST corpus 85
Figure 3.17 Summary of the process of corpus building 85
Figure 3.18 An example of *phrase type* query using PIE 99
Figure 3.19 An example of *wild card* query using PIE.....100
Figure 3.20 An example of the *lemma* results from PIE.....100
Figure 4.1 Mean of scores in the cloze test by 4 groups of participants and four-type head verb frequency 121
Figure 5.1 Mean of scores in the cloze test by 3 groups of participants (year of study) and 4-type head verb frequency 137

List of Appendices

Appendix A : NST approval letter .....
Appendix B : Malay Chair Grant offer letter .....
Appendix C : Ethics approval letter .....
Appendix D : List of private schools .....
Appendix E : Consent forms and information sheets.....
Appendix F : Cloze Test 1 .....
Appendix G : Cloze Test 2.....

CHAPTER ONE
INTRODUCTION

1.1 Background

The research to be reported in this thesis examines the acquisition of restricted collocations (RCs) by speakers of Malaysian English. This chapter briefly outlines the background of both the nature of Malaysian English and the nature of restricted collocations, and poses the central research questions to be investigated. The aims and a brief outline of the methods which will be used to investigate the research questions are discussed.

The English used in Malaysia is one of the varieties of New Englishes, and has emerged due to the spread of English around the world (Platt, et al., 1983; Pillai, 2006). English in the colonial and post-colonial era has reflected two different paradigms in the Malaysian context. A new policy which was established in 1967 promoted Malay as the national language. Platt et al. (1983: 5) suggested that this new language policy would encourage one to really talk of two kinds of Malaysian English. The first kind is still spoken by English-medium educated older Malaysians and some younger Malaysians of Chinese Indian ethnicity. The second one is that which is considered 'new' Malaysian English and is spoken by the younger Malay-medium educated Malaysians (Platt et al., 1983: 5). Regardless of the mode as first or second language, English is likely to continue to be important for Malaysians in world interactions. Over years of playing its various roles as

a language of interaction, there has emerged a variety of English that is distinctively Malaysian (Asmah Haji Omar, 1992).

Platt and Weber (1980) have established three levels of proximity to the English spoken by native speakers. Platt and Weber proposed that a lectal continuum existed from basilect, the sub-variety at the bottom of the continuum to acrolect, the variety at the top of the continuum, and mesolect, the variety in between. The acrolect is taken to be an idealized Standard British English and has been labeled by Ho and Platt (1993:12) as 'a prescriptive native speaker standard'. In similar vein, Malaysian English is perceived in terms of these various 'lects' (Baskaran, 1994: 25). According to Baskaran (1994) those are the three distinguishable sub varieties of the Malaysian English continuum, namely the acrolect (standard ME), mesolect (colloquial) and basilect (broken) varieties. Malaysian English is widely referred to as the colloquial variety spoken by Malaysians, but according to Gaudart (1997), it actually represents all the sub-varieties of English used by Malaysians. Morais (1997) claims that by placing Malaysian English on this continuum it is suggested that there is no neat division between the three sub-varieties. So, speakers of Malaysian English can be more or less acrolectal or mesolectal, depending on a combination of factors. Those underlying factors are decided based on whether it is a user's first, second or other language, the user's levels of proficiency and the context of use (e.g. the purpose, degree of formality, etc). Another factor is the socioeconomic and educational background of the user (Gaudart, 1997: 47)

The dialectal character of English used in Malaysia is still a subject of debate. Bao (2003) argued that Platt's (1975) and Platt & Weber's (1980) lectal continuum might be rather artificial and not sensitive to context. Bao sees this variation instead as the high to low levels of a diglossia, recognizing two distinct and independent varieties; the vernacular, informal English as the 'L' variety, and the standard, formal English as the 'H' variety. However, Fraser Gupta (1998) claims that those two approaches, the lectal continuum and the diaglossia approaches, are not necessarily in opposition as the former one focuses on the native speakers while the other focuses on the behaviour of all speakers in a formal context. According to Bao (2006), Malaysia and Singapore constitute a potential site for an interesting case study, as they are conditioned by different postcolonial experiences and have been through distinct developmental paths (Schneider, 2003). In Singapore, the government adopted an English-centered language policy. Therefore English is regarded as the administrative language as well as a medium of instruction in schools. Due to this an English diglossia has emerged (Bao, 2006). But in the case of Malaysia, Malay is the national language and standard English remains the language of an elite.

Baskaran (2002) points out that English is now adopted and adapted in the linguistic ecology of Malaysia, and all Malaysians should be proud of it with all its local 'nuances and innuendos'. Malaysian English today is 'a rich tapestry of a typical transplanted variety of English'. Malaysian English (ME) is one of the new varieties of English, with some distinct features including the localized vocabulary, pronunciation and grammar as well as pragmatic features (Pillai, 2006; Pillai and Fauziah, 2006, p.39). These distinct

features are also described in Baskaran (1987; 1994), Nair-Venugopal (2001), Platt and Weber (1980) and Platt et al. (1983).

Given the above, the present study embarks on a study of vocabulary acquisition. In particular, it examines the English collocations known by speakers of Malaysian English. The motivation for conducting this study is to explore the vocabulary knowledge of speakers of Malaysian English as it is assumed that non-native speakers of standard English do not share similar advantages to native speakers. It is due the fact that non-native speakers, particularly adult learners, are normally expected to acquire words rather than phrases (Kuiper, Columbus & Schmitt, 2009). In addition to that, Wray (2002) claims that non-native speakers acquire individual words separately which later pair for correct collocations.

The above notion has call for an urge to examine the lexical collocations acquired by Malaysian learners with exposure to both Malaysian English and New Zealand English. The study is restricted to Verb-Noun collocations of written English. The objective of the study is to assess the influence of exposure to both Malaysian English and New Zealand English on the acquisition of English restricted collocations through the educational life span.

1.2 The importance of the study

In this section I will indicate why this research is worth doing. I will list several studies conducted on collocations in second language learning in Asian countries as well as other

countries as matter of highlighting that these types of studies are limited and disregarded especially in Malaysia. I will conclude by proposing that the findings of the research may highlight a new dimension of treatment to collocational learning in Malaysia, as well as looking at the impact of malformed collocations (Chapter 6).

While vocabulary acquisition and collocation acquisition are considered vital, they have been largely disregarded in Malaysia, as shown by the fact that only limited research has been done in this area. The research done by Bayatee (2007) examined the cohesion and coherence in narrative and argumentative English essays written by 14 Malaysian and 14 Thai second year medical students at the National University of Malaysia in Kuala Lumpur and at Prince of Songkla University, Hatyai, Thailand. This quantitative and qualitative mixed type of study has revealed that both Malaysian and Thai medical students used more syntactic ties (reference and conjunction) than semantic ties (reiteration and collocations). Su'ad Awab (1994) has studied the collocational patterns of research proposals, while Nair-Venugopal (2001) has examined the unconventional collocations of English words, such as '*to shake legs'*. A study by Kamariah Yunus and Su'ad Awab (2011) highlights the collocational competence among law undergraduates who are studying at a local university in Malaysia. Ooi's (2000) study of collocational patterns of Malaysian-Singaporean English has revealed the influence of Asian values as several distinctive collocations have been identified by which Asians are likely to put stress on order and harmony. And most recently , a study on phrasal verbs (PVs) among Malaysian learners of English was done by Rafidah Kamarudin (2013) was done to examine the level of understanding and use of PVs. However, the methodology was quite

different with the present study as Rafidah was investigating survey or questionnaire apart from corpus work. Teachers' and learners' feedback were used for data collection. The corpus work was based on an existing corpus, English of Malaysian Students (EMAS). The overall research was looking at the understanding, perception of PVs, problems faced by learners, and how PVs were used in teaching materials.

By contrast numerous studies have been conducted on how helpful the knowledge of collocations (and formulaic language) in second language learning in other countries. Several studies have been done in Japan (Koya, 2004, 2005, 2006) looking at the acquisition of English collocations by Japanese learners. Miyakoshi (2009) conducted a study specifically on ESL learners' collocations. Her study focused on the verb-noun collocations by Japanese learners of English. In 1997, Lombard studied non-native speaker collocations by looking at the writing of native speakers of Mandarin. Bahn and Eldaw (1993) conducted an experiment consisting of a translation task and a gap-filling task with advanced learners of English who had German as a native language. Granger (1998) analysed the written performance of Advanced French students and found that learners overused very frequent collocations but underused creative constructions. Biskup (1992) collected interference errors made by Polish and German learners of English. There are more evidences of recent research been done in the area of formulaic language in general. Studies by Laufer and Waldman (2011) concerning verb-noun collocations, Siyanova and Schmitt (2007) on phrasal and prepositional verbs, and Millar (2011) on the impact of malformed collocations are among the related studies. Lindstromberg and Boers (2008a, 2008b) examined the mnemonic benefits of drawing learners' attention to

sound repetition commonly manifested in formulaic sequences, namely, alliteration and assonance.

Although English has had the status of a second language in the Malaysian education system for decades, many Malaysian learners are regarded as error prone in their use of English (Marlyna Maros, Tan & Salehuddin (2007); Saadiyah Darus & Kaladevi, 2009). Consequently, it is worth investigating factors causing those 'errors' or non-native-like expressions made by learners. On the same consequence, research by Ang et al. (2011) highlighted that the acquisition of phrasal vocabulary, specifically verb-phrase restricted collocations can be assessed using an error analysis approach, assessing 'errors' made by learners. However, for the analysis they applied standard British English norms. Yet, what is ideal is that the norms for second language acquisition be a combination of aspirational norms, those that the person learning the language aspires to, and actual norms, those of a target speaker community (K. Kuiper, personal communication, August 2, 2013). Aspirational norms can also be those of teachers, i.e. the norms that they wish their students to aspire to. Yet, all aspirational norms are value laden. As such they can be judged as to how realistic they are.

The above survey has suggested that the proposed research will make a useful contribution to the limited research done on the acquisition of restricted collocations in the Malaysian context. The researcher feels that by looking at the collocational patterns of Malaysian students' performance, the features and patterns of learners' collocations can been identified.

For this study, investigating collocational acquisition in English language learning is important as such study may inform us on the use of collocations in English language teaching in Malaysian context and the local school syllabus. It is intended that the findings of the research may provide the knowledge of collocations used locally. It will do this by identifying the patterns of collocations used by the Malaysian English learners as well as the exposure effect on the acquisition of collocations. Thus, this study may lead to a better understanding of the nature of acquisition of collocational patterns of written Malaysian English. Furthermore, the findings will shed light on the local norms for second language acquisition. This can be achieved by analyzing the verb frequency list extracted from the Malaysian English corpus and the BNC, as well as investigating the learners' scores of the cloze testing scores. This study will also suggest that relativised norms are more realistic. A special chapter will discuss these norms further, and a new approach to assess the non-native like responses is suggested. In Chapter 6 the coding for the non-native like responses are labeled as 'semantically plausible answers' which are coded using a novel approach and seem to be more realistic. This is how one might go about assessing them within the context of Malaysian second language learners learning English. Apart from that, the cloze test instrument devised for this study and the corpus which has been developed for it should also prove useful tools in assessing 'errors' or non-native like restricted collocations of Malaysian learners.

1.3 Aims of the study

The two studies which have been undertaken in this work adopt a model of lexical access for phrasal lexical items, namely superlemma theory (Sprenger et al., 2006; Kuiper et al.,

2007). This theory along with other relevant theories by Cutting and Bock (1997) and Sprenger, Levelt and Kempen (2006) look at how phrasal lexical items are stored and retrieved as well as looking at what is acquired. This theoretical framework is necessary in explaining how retrieval from the mental lexicon takes place in cloze tests. This model is discussed in Section 2.9 of the thesis.

The present research aims to address the following leading questions:
The first research question addresses the acquisition of restricted collocations of Malaysian English among learners in Malaysia. This research question will be addressed in Chapter 4. The main objective of this study is to gauge the influence of exposure to Malaysian English on the acquisition of restricted collocations through the educational life span. Researchers have different opinions on how collocations are best learnt. Schmitt (2000, pp.116) has raised the issue of how language learners are able to acquire thousands of word families. Schmitt argued that this amount is probably too large to be learnt solely from formal study, so LI vocabulary knowledge is best acquired implicitly, through extensive exposure to the target language. This question and its consequential hypothesis can be framed as 1.

1. How does the duration of exposure to the collocations of written Malaysian English affect the acquisition of Malaysian English collocations through the educational lifespan of learners of Malaysian English?

The hypothesis to be investigated in answer to this question is that acquisition is positively correlated with exposure to RCs as measured by the length of time a learner has been exposed to the target language.

The second research question concerns the acquisition of restricted collocations of standard English of Malaysian learners abroad, specifically those who are studying in New Zealand. The main objective of the second study is to see the influence of exposure to standard English during the respondents' undergraduate studies. The learners' knowledge of restricted collocations is examined to see the influence of exposure to New Zealand English, as it is suggested by Randall (1980), Schmitt (1998) and Sökmen (1993) that learners will acquire more 'native-like' phrasal vocabulary as their proficiency increases and proficiency can only increase as a result of exposure to the target language over time, i.e. it is exposure graded. While this may be the case with many aspects of acquisition, it is an open question whether it is also the case with the acquisition of restricted collocations by international students. This study will be discussed in Chapter 5. The relevant research question and hypothesis is 2.

2. How does the duration of exposure to New Zealand English affect the acquisition of standard English collocations in Malaysian international students?

The hypothesis to be investigated in answer to this question is that the acquisition of RCs is positively correlated with exposure as measured by the length of time a learner has been exposed to the target language.

It is assumed that the frequency of a lexical item in corpora is a factor in acquisition because the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to the item, resulting in research questions 3 and 4.

3. What role does the frequency of a clozed head verb of a verb phrase collocation, as a measure of the likelihood of having been exposed to a VP collocation, play in both the case of 1 and 2?

The hypothesis to be investigated in answer to this question is that the frequency of the head verb in corpora of the target language is positively correlated with acquisition.

4. What role does the frequency of a restricted collocation, as a measure of the likelihood of having been exposed to a VP collocation, play on its acquisition in the case of 1 and 2?

The hypothesis to be investigated in answer to this question is that the frequency of the RC in corpora of the target language is positively correlated with acquisition.

By analyzing the responses of some of the cloze tests in more depth, the research also aims to answer question 5.

5. Do learners proceed their strategies by moving through a stage of using a literally plausible but unidiomatic phrasing through to the use of restricted collocations?

The hypothesis to be investigated in answer to this question is that learners do pass through a stage where, in individual cases, they fill a cloze gap with a guess which is not

semantically plausible, i.e. the learner does not understand the clause in which the cloze item is found. They then are able to fill the cloze gap with a semantically plausible item and lastly, they are able to select the idiomatic filler for the gap. Furthermore the choice of fillers will be sensitive to the frequency of the filler in the relevant corpus.

Perhaps the most obvious reason to explain as of why written English has been selected as the target for this study due to the fact that lexical phrases are extensively used in written texts and definitely this knowledge of vocabulary is a prerequisite for writing (J. Li and Schmitt, 2009). Thus, this provides evidence that lexical phrasal play great importance for L2 writers. Another reason is that it is the educational target for learners. It may be supposed that all the learners are already proficient in spoken Malaysian English of various genres. It explains why the corpus used for Study 1 is in standard written Malaysian English and that is what the educational system effectively has as its aspirational norm for students. Also, it is the source of the verb frequencies as well as the restricted collocations used for the test. In that regard, the use of vernacular story with the selected restricted collocations appropriately embedded is for the sake of eliciting cloze responses from a form that is attractive enough for respondents to want to engage with it. If there was a vernacular Malaysian English corpus or even vernacular source of vernacular restricted collocations they might be chosen for the test. It is therefore considered that the restricted collocations are common to both standard written Malaysian English and vernacular Malaysian English.

As can be seen from the above research questions one of the methods which allows for the investigation of the retrieval of phrasal expressions is cloze testing (Kuiper, Columbus & Schmitt, 2009). Cloze testing is considered one of the best suitable tools in assessing language ability, namely second language proficiency and reading comprehension skills (Alderson, 1979; Abraham & Chapelle, 1992; Dörnyei & Katona, 1993; Kobayashi, 2002). The cloze tests designed for the two studies consist of twenty cloze gaps in a vernacular narrative to be filled by the participants. The results were analyzed in two different binary methods. The first binary analysis supposes that the selection of the cloze item is either 'right' or 'wrong', 'native like' or 'non-native like'. The results of this analysis are presented in Chapters 4 and 5. The second analysis is in terms of the respondent gap-fills in terms of providing:

i) an idiomatic response

ii) a semantically plausible response

The later analysis is undertaken on the responses in the second study. The details of this analysis are reported in Chapter 6. This chapter deals with a fifth research question, namely how exactly does the experience of studying abroad impact the sophistication of word knowledge and lexical organisation. The immersion period is valued for its capacity for extensive exposure and is believed to elevate the progress of vocabulary acquisition (Fitzpatrick, 2012). This exploratory study is tested through the frequency of the syntactic heads of restricted collocations in the mastery of such items by learners of English. Kuiper, Columbus and Schmitt (2009) suggest there exist frequency effects on acquisition.

A number of tools are required for these studies: the main ones are suitable cloze tests and relevant corpora for assessing frequency of the target items. The first study, since it deals with the acquisition of restricted collocations in written Malaysian English requires the creation of a new corpus, the News Straits Times (NST) corpus, as there is currently no large corpus of written Malaysian English which offers a satisfactory representation of the lexicon of standard Malaysian English. In fact, the existing Malaysian English databases have serious disadvantages not just for the limited size but also because they do not form a balanced corpus of Malaysian English. What was observed was that they were compiled for small language projects and may consist of several students' essays and few excerpts of news reports as in the case of the ICE corpus of Malaysian project. This corpus is still being developed and is untagged.

To this end a large corpus of 19 million words was created to check for word and collocation frequencies, since frequency as a measure of exposure and acquisition appear to be linked. The frequency list secured from the NST corpus contributed to the development of a cloze testing item for the study. The cloze test was later used to measure the acquisition of restricted collocations of written Malaysian English in Chapter 4. A specific discussion regarding the corpus development is presented in Chapter 3.

The line of argument running through the study and the hypotheses that will be tested are as follows:

Exposure grading

Study 1(which will be discussed in Chapter 4): The exposure to English in Malaysia over the education life course of students in Malaysia will result in increased acquisition of standard local English restricted collocations. Specifically:

i. Learners with more months of exposure to English will perform significantly better than those with fewer months of exposure;
ii. Learners at higher study levels will perform significantly better than those from lower study levels (year of study);
iii. The acquisition of restricted collocations is age-graded, which in this sense age is a proxy for numbers of months/ years of exposure to a language;
iv. The frequency of usage of head-verbs contained in verb plus complement formulaic sequences is tightly linked to the acquisition of such sequences.

Study 2 (which will be discussed in Chapter 5): The exposure to New Zealand English will result in increased acquisition of native-like restricted collocations by international students from Malaysia during three years of study in New Zealand

i. Learners with more months of exposure to English will perform significantly better than those with fewer months of exposure;
ii. Learners at higher study levels will perform significantly better than those from lower study levels (year of study);

iii. The acquisition of restricted collocations is age-graded, which in this sense age is a proxy for numbers of months/ years of exposure to a language;

iv. The frequency of usage of head-verbs contained in verb plus complement formulaic sequences is tightly linked to the acquisition of such sequences.

Study 3 (which will be discussed in Chapter 6): The relationship between the exposure to a lexical item and the learning of that item. The hypotheses to be investigated here are that:

1. Individuals who have the higher scores of idiomatic responses will also have higher rates of plausibility for their non-idiomatic responses;

2. Of the non-idiomatic responses more will be at high end of the frequency spectrum.

1.4 The shape of the thesis

Overall the thesis has seven chapters discussing the acquisition of English collocations by learners from Malaysia. Chapter 2 provides general theoretical background of the notion of the acquisition of restricted collocations and how exposure and frequency are closely linked to this phenomenon. Chapter 3 discusses the development and establishment of the new corpus of Malaysian English which has been used for generating the word frequency list for testing the acquisition of vocabulary items. This is a significantly large corpus and can be considered as a beginning of a balanced corpus. Chapters 4 and 5 look at how restricted collocations are acquired using a quantitative approach. Chapter 6 is an

exploratory study exploring learners' selectional patterns favouring particular heads of phrases. It is supposed that the production task of filling a cloze gap requires learners to fill the gaps of the testing items with either restricted collocations or with semantically plausible verbs. The task of supplying the missing verbs leads to retrieving them from the mental lexicon and it is assumed that context and some of the constituents can activate the missing verbs from the mental lexicon (Jackendoff, 1995). This chapter will present a case study where learners' preferences of head verbs may reflect learners' exposure to particular restricted collocations.

Having introduced my general research aims, I will conclude the findings of Study 1 and Study 2 (presented in Chapter 4 and 5) in Chapter 7. The research in these two chapters has focused on groups and the results are cohort results. However, a new constructive tool used in Chapter 6 has outlined an approach to acquisition which is focused on individual acquisition. I will conclude as well by proposing that the semantic plausibility metric which is used as a tool for the study in Chapter 6 has the potential to be used as a measure of vocabulary acquisition. So, in Chapter 7, I will present my overall conclusions and summarize the whole thesis, as well as providing suggestions for future research.

CHAPTER TWO
REVIEW OF THE RELEVANT LITERATURE

2.1 Introduction

This section outlines the background of the nature of restricted collocations and theories related to the acquisition of restricted collocations, and later associate this notion with the effects of frequency and exposure. I will also outline the theoretical background and look at the theories related to the studies.

2.2 Phraseology

Despite of the fact that 'Phraseology is a fuzzy part of a language' (Altenberg, 1998:101), special interest has been paid to the field of phraseology (Cowie,1998) for a number of years. This is reflected in several works by prominent linguists, for instance Kuiper and Austin (1990), Pawley and Syder (1983), Becker (1975), Bolinger (1976), Coulmas (1979), Nattinger and Decarrico (1992), Wray (2002), and Schmitt and Carter (2004). The domain of phraseology is variously named due to 'lack of common descriptive approach and no consensus on the analytical procedures to be followed' (Howarth, 1996, 6). Although there is no generally-agreed overall term to describe the whole spectrum, there is a sign of increasing acceptance of 'phraseology' as a convenient name for this field of study, and it is seen as a broad and neutral term to describe the phenomenon. Thus, the expressions like 'word combinations' (Cowie, 1994; Howarth, 1996), 'phrasal lexemes' (Moon, 1998), 'phrasal lexical units' (Glaser, 1998), and 'phrasal lexical items' (PLIs) (Kuiper, 2009) are given and this is significant evidence that the focus is on the same area of interest. The importance of phraseology is being highlighted and it is

becoming increasingly clear that it is an important element of language learning and use because both written and spoken discourse contain large percentages of formulaic language (Schmitt, 2005). This notion is supported by Erman and Warren's (2000) calculation that 52-58% of the language they analyzed was formulaic. Apart from that, Foster (2001) estimated that 32% of the texts studied were found to be formulaic language. Moreover, Sinclair (1991) argues that the dominant structure feature of language is the *idiom principle*, rather than the rule-based *open-choice principle*. This model will be discussed further in Section 2.7 of the thesis.

In 1997, Jackendoff examined naturally occurring English data from the TV show 'Wheel of Fortune' (WoF). It showed a significantly high frequency of collocations, idioms and 'prefabricated' phrases. The WoF corpus has a high percentage of both phrases (types) and frequency of phrases (tokens). This leads to the suggestion that people may know at least as many formulaic sequences as single words (Jackendoff, 1995). Likewise, formulaic language is used to fulfill various functions in language use (Nattinger and DeCarrico, 1992). These include such (speech) acts as apologizing, giving directions and making requests, and they function as a quick and reliable way to achieve the related speech act.

Apart from the fact that language includes a significantly a large amount of formulaic language serving various functions, these formulas help proficient speakers to become fluent speakers (Schmitt, 2005). In addition, formulas are believed to help speakers deal

with the demands of real-time language production and comprehension while maintaining fluency (Coulmas, 1981; Kuiper, 2004; Pawley & Syder, 1983).

Kuiper (2004) suggests that the use of formulaic speech supports fluency. The review which was based on Lord and Parry's work has unveiled how Yugoslavian illiterate oral bards composed their poems in real time yet at the same time maintained fluency. This difficult task required them to face a mobile audience as well as to react to the performance. The key was that they used formulaic phrases and followed known outlined plots, and this can only be achieved through exposure. Although the songs had been performed before, they were never performed exactly the same way twice. So, their unique oral performances reflect how tradition was acquired. Kuiper has drawn a significant conclusion from Lord and Parry's work that it was a theory of language acquisition for formulaic varieties of speech. The consequence is that formulaic performance takes place where speakers are under pressure. The pressure involved is specifically the pressure on poets' working memory. There are several similar studies investigating speech production which look at high-pressure but routine situations, including sports commentary (Kuiper and Austin, 1990) and auctioneering (Kuiper and Haggo, 1984).

According to Nation (2001), in order to develop fluency, all collocational sequences are deemed important and learners should encounter these sequences repeatedly. In terms of native speakers, Pawley and Syder (1983) suggest that though native speakers have cognitive limitations in how quickly they can process language, they are able to produce

language seemingly beyond these limitations. Besides, speakers seldom pause in the middle of a clause and they can fluently say multi-clause utterances. It is presumed that the formulaic varieties of a language can be fluently produced because they are already memorized as prefabricated phrases. The phrases are stored as single wholes and are instantly available for use without the cognitive load of having to assemble them on-line as one speaks (Pawley and Syder, 1983; Kuiper and Haggo, 1984; Kuiper, 1996). Pawley and Syder further suggest that our mind uses its vast memory to store these fabricated phrases in order to compensate for limited working memory.

Becker (1975) raises the issue that the understanding of memorized phrases, or 'ready-made lexical units' or 'prefabricated' lexical units to borrow the term use by Bolinger (1976) is actually basic to the understanding of language as a whole. In his argument, Becker proposed six major categories of lexical phrases. Those categories are; polywords, phrasal constraints, meta-messages, sentence builders, situational utterances and verbatim texts. The current terms being used today may diverge from Becker's, yet these subclasses are explicitly laid down and have been a handy guideline for researchers, linguists and language users in general. The above research has suggested that lexical phrases are real, and there is the possibility that individuals might know more lexical phrases that single words (Becker, 1975).

Nattinger and DeCarrico (1992) adapted Becker's classification by proposing a smaller number of classifications which are categorized as polywords, institutionalized expressions, phrasal constraints and sentence builders. They stressed the need for more

empirical fieldwork to be done as learners' phraseological skills are limited, or, in other words, they use a very limited number of native-like prefabs and too many foreign-sounding prefabs (Granger, 1998).

Furthermore, Howarth (1996) presents a very significant study on the use of prefabricated language in the production of native and non-native writers of English. A framework is developed which focuses on 'restricted collocations'. Some deviations are found in the writing of advanced foreign language learners from that of native academic writing, and this is regarded as being due to lack of knowledge of what is conventional in the use of academic, field-related collocations. The corpus of advanced learner writing which consists of academic essays produced by overseas postgraduate students shows some collocational errors made by the students. It suggests that learners do not approach the phenomenon from the same direction as native speakers. And it is found that the existing published teaching materials offer very little help as they fail to recognize the nature of collocations.

Granger (1998) provides comparative analyses of word-combinations in the writing of native speakers and foreign learners of English. The aim is to identify the norms implicitly recognized by native speakers and to demonstrate and explain how the usage of foreign students departs from them. The study is concerned with the analysis of phraseology in the written English of advanced foreign students. The author is interested in identifying the phraseological norms implicitly recognized by native speakers. This study explains how the usage of foreign learners deviates from such standard norms. She

compares native and non-native varieties of English with the hypothesis that learners will make less use of prefabricated language (collocations and formulae) than native speakers.

Overall, the above research illustrates that formulaic language does not just facilitate social interaction, its use is also closely related to fluency and language functions. It summarizes that formulaic sequences are an important element of language use.
We turn now to the vocabulary for phrasal lexical items itself (Wray, 2002) i.e. *chunks, formulas, multiword units, collocations, etc.*, and types. All these really depend on the degree of fixedness, institutionalization/ conventionality and opacity/ non-compositionality (Moon, 1997, p.44). This is reflected in Moon's (1998) three macro-categories of phrasal lexemes: anomalous collocations, formulae and metaphors. Her study reveals that many phrasal lexemes are frozen in particular transformations, such as the passive. Moon's established typology is simply a way of classifying a wide range of various types of 'fixed expressions'.

The notion that lexical items are categorized into two subcategories is clearly discussed in Kuiper (2009). There are structurally simple lexical items and structurally complex lexical items. The structurally simple items are monomorphemic, where they have only one meaning-bearing element. Word structure and syntactic structure are the two elements which determine the structure of the structurally complex lexical items. Kuiper proposes that the syntactically complex items are phrasal lexical items (PLIs) (Kuiper, 2009).

Despite of the various terms given, researchers are investigating the same phenomenon: the study of word combinations (Howarth, 1998). The focus of this present research is restricted to the aspect of phraseology that is significant to the needs of second language learners of English, in particular their use of collocations. This study is chiefly concerns restricted collocations.

2.3 What is collocation?

Collocation has come under the spotlight as one particular area of vocabulary research which has been seriously discussed from the standpoint of language use and language learning for the last few decades. The initially most influential work done by Palmer (1930s) and Firth (1957) has motivated other scholars to explore the phenomenon further. In relation to that, many linguists have focused not on each word in a sentence, but on the combination of words in terms of productivity (Wray, 2002; Moon, 1997, 1998; Koya, 2005; Howarth, 1996, 1998).

According to Firth (1957), collocations play an important role in a word's meaning, and are defined as follows:

> The statement of meaning by collocation and various collocabilities does not involve the definition of word-meaning by means of further sentences in shifted terms. Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course,

> collocation with *night*. This kind of mutuality may be paralleled in most languages. (p. 196)

Firth highlights the point that collocation refers to co-occurring associations of two or more lexemes in a specific range of grammatical constructions, and that mutual expectancy of words is a distinguishing feature of collocations. Other researchers who agree on similar notions are Sinclair (1966, 1991), Bollinger and Sears (1968) and Carter (1987).

The following definitions of collocations share a similar attribute; they involve at least two or more words. For example, Sinclair (1991: 170) defines collocation as 'the occurrence of two or more words within a short space of each other in a text.' Bollinger and Sears (1968) point out that collocations are normally acquired relatively late through the process of L1 vocabulary acquisition. Carter (1987) has described collocations as a group of words which co-occur repeatedly. Jackson (1988) uses the term collocation to refer to the combination of words that have a certain 'mutual expectancy', and his definition therefore excludes clichés which are regarded as 'ossified collocations'. Fellbaum (2007) defines collocations as lexical entities consisting of words that tend to be found together. The words are thus associated strongly enough to qualify them as 'fixed expressions'.

The term *collocation* is not yet fixed (Miyakosyi, 2009) and is confusing (Moon, 1998, p. 26). A number of labels have been given to the units of formulaic language, including

phraseological units (Gläser, 1986), word combinations (Cowie, 1994), and phrasal lexemes (Moon, 1998). Despite differing labels, researches are, more or less, investigating the same phenomenon: the role of patterning other than grammatical patterning of words and phrases in communication.

The present research will adopt Kuiper's (2009) term 'phrasal lexical item' as his comprehensive approach best defines and describes the language phenomena that the researcher will be studying. According to Kuiper, 'each PLIs is a lexical item with its own entry in the mental lexicon of a speaker who knows it.' (p. 18). All PLIs are restricted collocations. Restricted collocations are phrasal lexical items where the grammar and the semantics of an expression can have a number of words which function in a position but only a subset is lexicalized. In this sense, the term PLIs is considered as a superordinate term which is given to describe similar phenomenon of words combination e.g. phraseological units (Gläser, 1986), and phrasal lexemes (Moon, 1998).

2.3.1The importance of collocations in language learning

"People do not speak in words, they speak in phrasemes" (Mel'čuk, 1995: 169).

Corpus research has shown that the presence of phrasal vocabulary is ubiquitous (Nattinger and De Carrico, 1992: 66; Kuiper, Columbus & Schmitt, 2009) in both the spoken and written discourse of native speakers (Altenberg, 1998; Howarth, 1998 Kuiper, 2004). As regards the size of the phrasal lexicon, Kuiper (2009) has suggested there might possibly be a larger number of PLIs than single word lexical items (Kuiper,

Columbus & Schmitt, 2009) in the lexicon of a native speaker. Pawley and Syder (1983: 213) claim that the number of 'sentence-length expressions familiar to ordinary, mature English speaker probably amounts, at least, to several hundreds of thousands'. In addition to that, Benson, Benson and Ilson's (1986) collocational dictionary contains over 90,000 entries. Jackendoff (1995) states that the number of fixed expressions stored in the mental lexicons are at least of the same order as the number of single words. Not withstanding such estimates, it is hard to gain a reliable idea of the size of the PLIs in the mental lexicon of individual speakers (Newsome, 2005; Kuiper, Columbus & Schmitt, 2009).

Wray (2002) has outlined few functions of formulaicity where these sequences appear able to enhance fluency of speaker's output (by reducing the processing load), though the case of learners might not match the native speakers. However, the processing advantage of formulaic language can also be exploited by learners to increase fluency. Another function is that speakers are able to express their identity as an individual as well as a member of a group. In addition to those functions, J. Li and Schmitt (2009) have suggested that the result of learner frequent use of these sequences would become defining markers of fluent writing thus later would fit the expectations of readers in academia. Also, these sequences make the learners' task easier due to that they are dealing with the ready-made sets of words.

Given the significant functions of PLIs, it is not surprising that they are regarded as increasingly important in both first and second language vocabulary acquisition in all languages, as all languages appear to have PLIs. The acquisition of these PLIs is a

significant part of vocabulary acquisition for native speakers. It is therefore significant for the acquisition of English by native speakers and non native of English. Consequently collocation has become an important area of vocabulary research (Koya, 2005) as well as language acquisition research. Howarth (1996) highlights the importance of collocation for second language learners as follows:

> They can be considered most centrally involved in the process of composition at clause level, therefore potentially a sensitive indicator of learners' acquisition, … (p. 24).

Howarth (1998) found that imperfect control of idioms and collocations can have a significant effect on the effectiveness of student writing which may not meet the stylistic expectations of the academic community. He suggested that in fulfilling the very demanding requirements of academic assessment, even advanced non-native writers tend to fail to communicate effectively their understanding of the subject matter due to incomplete linguistic competence.

Other researchers (Bahn & Eldaw, 1993; Bahn, 1993; Channell, 1981; Willis, 1990) have discussed the importance of collocations, and they agree that learners must master how words combine or collocate with each other in order to develop their vocabulary proficiency.

One of the ways in which the study of collocations has come to the fore in the last few decades has been through corpus-based research. Corpus-based research on collocations is not new. A British linguist, H.E. Palmer (1933) undertook corpus-based research on recurrent combinations of English words which led him to the conclusion that there were unrecognized patterns of word use in everyday language which hardly fit into either of the traditional categories of lexis or grammar. This research has motivated and inspired other researchers (Crystal, 1987; Halliday, 1966), particularly specialized in computer-assisted analysis, to carry out more research in this area. Later those unrecognized patterns were named collocations.

The term *collocation* was first introduced by Firth in 1957 (Koya, 2005) and it is known as a 'Firthian' term (Kjellmer, 1982: 25; Fernando, 1996: 29; Nation, 2001: 317). It is defined here as 'actual words in habitual company', e.g. *pouring rain*, or in statistically significant proximity. In relation to this, Benson (1990) viewed a collocation as an arbitrary and recurrent word combination. Smadja (1993) believes that collocations are difficult to translate across languages, or even across different dialects or varieties of English. Ooi (2000) has extended this observation by suggesting that collocations or multi-word patterning of different single concepts represent or even express the reality and values inherent in a speech community.

While collocations have been widely recognised as an important part of vocabulary acquisition (Howarth, 1996, 1998; Sinclair, 1991; Nation, 2001), this finding has not been of great importance among educators and learners in Malaysia. Ideally, applied and

educational linguists should have emphasized the importance of drawing learners' attention to standardized collocations as there are numerous reasons why a command of these restricted collocations in L2 is believed to be beneficial to learners**.** Boers et al. (2006) and Stengers et al. (2010) highlight the fact that mastery of standardised phrases such as collocations not only aids learners to become proficient speakers as they attain a native-like command of the language (Pawley and Syder, 1983), but also facilitates learners' fluency under real-time conditions (Skehan, 1998; Wray, 2002)**.** On the other hand, while this knowledge of collocations has been recognized as an important aspect of language learning (Howarth, 1998; Nation, 2001), many studies have shown that mastery of collocations has created a difficulty for learners even at advanced study level (De Cock, 2004; Granger, 1998; Nesselhauf, 2003; Siyanova & Schmitt, 2007; Wray, 1999). Howarth (1998) reports that advanced non-native writers may fail effectively to communicate their understanding of a subject matter due to incomplete linguistic competence in its subject matter codes (Pawley, 1991) rather than because of academic weakness. Howarth's study analysed ten essays written by international students who were attending an MA course in the UK. Nine of them were teachers of English as a second or foreign language, while one taught German to English-speaking children. They came from eight countries; Botswana, Burkina Faso, Germany, Greece, Hong Kong, Japan, Taiwan and Thailand. Scarcella (1979) similarly found that the thirty advanced' ESL respondents in her study had difficulty acquiring common verbal routines. The cloze test instrument used in her study was also tested as a pilot study conducted with a group of adult ESL students. The test scores of these two groups were both found to be low, where the average score was 30%.

In this thesis similar research trajectories will be applied in a different language setting and context, Malaysia, where the English spoken is one of the varieties of New Englishes.

2.4 Restricted collocations

A number of definitions are given to describe restricted collocations. As Kuiper (2004: 51) highlights, restricted collocations appear in all types of speech and can be defined as *'pairs of words which occur together in ways that are more restrictive than the grammar of the language requires'*. Restricted collocations are not formulae as they are not restricted by anything accept for their meaning. Kuiper uses as examples *give offence* and *take offence*. The only 'acceptable' verbs used by native speakers of English are 'restricted' to these two verbs. It is impossible to use *donate offence* or *accept offence*.

Apart from Kuiper, Howarth (1996) describes research done on phraseological performance of non-native writers of English in academic writing, in which the findings are significant for the study of collocations. Howarth's definition of restricted collocation is as follows:

> 'combinations in which one component is used in its literal meaning, while the other is used in a specialized sense. The specialized meaning of one element can be figurative, delexical or in some way technical and is an important determinant of limited collocability at the other..'
>
> (Howarth,1996: 47)

In defining restricted collocation, Cowie (1991: 102) describes restricted collocations as 'word-combinations in which one element (usually the verb) has a technical sense, or a long established figurative sense which has since lost most of its analogical force'. He discusses few examples such as *run a deficit, abandon a principle,* or *champion a cause,* in which the object noun limits the choice of verb to only one or two. Moon (1998: 27) sees that this kind of collocation occurs where 'a word requires association with a member of a certain class or category of item'. She further proposes that they are semantically and lexicogrammatically restricted. Moon shares Aisenstadt's (1981) concept of 'restrictedness' by stating that a word contains a particular meaning only when it is collocated with certain other words. Aisenstadt (1981) refers to these occurrences as restricted collocations and provides the examples of *face the truth/ facts/ problems*.

In relation to that, Granger (1998) investigates restricted collocations which focus more on amplifiers functioning as modifiers of adjectives. The findings show 'sharp differences between native and non-native usage' (Cowie, 1998: 13). Her study finds that *completely* and *totally* are significantly overused by the learners. To summarize, restricted collocations involve preferential selection of word combinations where the combinations are arbitrary, and they might also be idiomatic (Kuiper, 2009).

The present study mainly focuses on verb-noun lexical collocations. In part this is because they are a significant subset of restricted collocations. Moon (1998) found that

verb phrase idioms are the most frequent fixed expressions in the Hector Corpus. Cowie (1992) also reports the percentage of verb phrase idioms and restricted collocations in news stories and feature articles to be around 40 percent. Verb-noun combinations are regarded as key combinations in producing clauses and sentences, and they are the most often selected in the previous empirical research (e.g. Bahn and Eldaw, 1993; Bahns, 1993; Biskup, 1992; Nesselhauf, 2003). These studies have suggested that more focus is to be placed on verb-noun collocations, since it is the verb that causes the greatest difficulties for learners.

The acquisition of verb-noun collocations (e.g. *make a mistake*) causes great difficulties to (adult) L2 learners due to several reasons (Boers, Demecheleer, Coxhead and Webb, 2014). It is typically found that learners tend to substitute the verb in the collocations by an unconventional choice such as e.g. *do a mistake*. A study by Laufer and Waldman (2011) found that there hardly any differences in productive knowledge of verb-noun collocations between lower and upper intermediate groups of EFL learners. And for the substitution case it is more likely due to the interference from the mother tongue (Yamashita and Jiang, 2010; Nesselhauf, 2005). Apart from that, learners may not possibly see the needs to attend to the verb in interpreting the phrase. Learners may find the verb contributes relatively little to the semantics of some collocations. The slow uptake of verb-noun collocations by learners may lie in the lack of distinctiveness of the verbs, where the verbs may be treated as synonyms by the learners.

Thus, what is needed most is multiple encounters with a verb-noun collocation are required for the learners in establishing a firm association between the particular verb and the particular noun.

2.5 The acquisition of phrasal vocabulary by native and non-native speakers

In discussing the acquisition of phrasal vocabulary of native speakers, Peters (1983) proposes that formulaic sequences play a highly significant role in the language acquisition of some children. An additional finding is that children who use a Gestalt strategy for language learning tend to make use of formulaic sequences regardless of the length or size of the units they can acquire. In fact, it has been found that quite lengthy strings of adults' formulae can be treated as a single unit by a child (Plunkett, 1993: 44).

It is assumed that there are two types of formulaic sequences in child language acquisition (Peters, 1983: 82; Hickey, 1993: 29; Wray, 2002: 106). The first is *underanalyzed strings* which are sequences which adults understand to have a more complex structure than a child does. The second one is referred as *fused strings*, utterances which the child first uses whole and then segments them into components. After that the child analyzes or 'recreates' and stores them whole. The later strategy is considered using *shortcuts* and requires little processing attention as the strings are holistic in nature.

Since children are exposed to many formulaic sequences in their input, a conflict may arise regarding what to analyze and what to store holistically. Wray (2002: 130) suggests that a 'needs-only analysis' is applied as children will simply analyze whichever strings

need analyzing for social communication. In this sense, children would prioritize analysis only when the need arises. In effect, many word strings would remain unanalyzed and be retained as single units, thus maintained into adulthood. This need-only analysis will later lead to dual storage, where simple units or individual lexical items and formulaic sequences are stored together.

Many studies have shown that formulaic sequences are important aspects of language and language use in second language (L2) studies (Kuiper, Columbus & Schmitt, 2009). In studies such as those undertaken by Spöttl and McCarthy (2004), Nesselhauf (2003), and Schmitt, Dornyei, Adolphs and Durrow (2004), learners are found to be able to produce formulaic sequences in tests designed for them. Spöttl and McCarthy (2004) revealed that 14 Austrian multilingual learners were highly capable of selecting the correct formulaic sequences for a context. These learners were tested using a multiple-choice test of formulaic sequences. Another similar finding regarding learners' successful reproduction of formulaic sequences was a study by Schmitt, Dornyei, Adolphs and Durrow (2004) where the participants, postgraduate students, were found to have the knowledge of 16.84 out of 20 formulaic sequences measured with a multiple-choice test.

However, apart from these good results of high performance by L2 learners, there is also evidence that formulaic sequences tend to lag behind and that their acquisition seems to be difficult for learners (Siyanova & Schmitt, 2007; Altenberg & Grenger, 2001). In fact, Laufer (2000) mentioned that learners often intentionally avoid using these forms. Alternatively, learners tend to opt for single word use as this is assumed to be an easier

escape from having to recall a phrasal lexical item. With respect to this, Wray (2002: 76) proposes that non-native speakers acquire individual words which are later paired for idiomatic collocations. This scenario shows that paying more attention to words than sequences might give learners more 'power' or sense of control over the second language they acquire.

The art of using formulaic sequences effectively is not a simple task for learners (Wray, 2002: 171). As De Cock (2000) also found, some formulaic sequences tend to be overused, underused or even simply misused by learners. Thus when L2 users write or speak in a target language, their performance may be perceived as non-native like due to their limited command of acceptable lexicalized phrases. They often confine themselves to a limited range of familiar vocabulary items or produce expressions that sound odd, unidiomatic, or unintentionally amusing, and may write less fluently and idiomatically than natives do (Miyakoshi, 2009). Pawley and Syder (1983) claim that L2 learners are not aware of the importance of formulae in a second language acquisition, and they at the same time do not have sufficient collocational competence. If there is insufficient exposure to formulae then they need to parse an incoming sentence. They parse an incoming sentence in order to find out what the sentence means. They probably know all of the words of the sentence and understand what that particular sentence means. And in that sentence there is a restricted collocation. What they frequently hear is *sit an exam*, and the verb *sit* is used. But they never hear *do an exam*, but there is no negative evidence (K. Kuiper, personal communication, 22 July, 2013) that native speakers do not say *do an exam*. How do they know that *do*, though a perfectly semantically plausible verb in this

position, is not the conventional verb used in this context? Since there is no negative evidence, this knowledge must be received from frequency effects. Yet, how many times would be sufficient enough?

Clearly it is difficult when the verb which is head of the restricted collocation is infrequent by itself. If learners tend to hear this verb in a restricted collocation and not much elsewhere, then it totally depends on the frequency with which they hear the restricted collocation. But we know that restricted collocations are, by comparison with the frequency of single words, infrequent.

2.6 The importance of frequency approach in lexical studies

This section will look at how and why corpus frequency is a significant matter in vocabulary acquisition. The discussion will also provide insights into the relationship between the frequency of collocations and their learnability. It will be suggested that frequency is a proxy measure for the likelihood of a learner being exposed to a vocabulary item, including a phrasal lexical item such as a collocation.

The methodology of using a corpus or corpora is now well established. The areas of research utilizing corpus-based methodology include lexicography, pragmatics, stylistics, psycholinguistics, as well as literary studies. A corpus-based approach can also be applied to many aspects of linguistic inquiry. Having a systematic approach to the study of lexical and grammatical collocations allows this method to be adopted by researchers, and linguists in general. Corpora provide us with large collections or databases of texts

from a language. Specifically, the *'insights from corpus research have revolutionized the way we view language, particularly words and their relationships with each other in context'* (Schmitt, 2000: 68). They include looking at the relationship between frequency and collocations. Large corpora, i.e. Nation's (1990) list, are required to make such a study possible, at the same time avoiding painstaking and tedious hours of manual labour. Yet, in using corpora for linguistics inquiry, we need to bear in mind the cautions voiced by Biber's (1989) study of the difference between written and spoken corpora. When using corpora to investigate written language it is also necessary to bear in mind the possibility that restricted collocations are more likely to be used in spoken language than written language (Shin, 2007).

It is impossible to judge how many words individual people are exposed to as there are no records of personal corpora, i.e. corpora 'in the head' containing everything that an individual has heard or acquired. So, the only accessible corpora are the ones containing texts of a more general kind, i.e. text corpora. Thus, in this sense corpus frequency is taken as a proxy for the probability that a language learner has been exposed to a lexical item. Frequent collocations will therefore be the most useful because '*frequent collocations have greater chances of being met and used'* (Shin and Nation, 2008). Shin and Nation also found that *'the shorter the collocation, the greater the frequency'.* Their study revealed that two-word collocations make up 77 percent of the total number of collocations in the spoken section of the British National Corpus (BNC).

We might thus ponder how important collocations are in classroom learning? It is not simply because they exist that they deserve teachers' and learners' attention. The justification is that if the frequency of a collocation is high and it occurs in many different uses of the language, then it deserves attention (Nation, 2001). Added to that, frequent collocations deserve teachers' and learners' attention if their frequency is equal to or higher than other high-frequency words (p. 325).

Vocabulary acquisition is generally known to be sensitive to the frequency of vocabulary items (Ellis, 2002; Kuiper, Columbus & Schmitt, 2009; Read, 1988; Schmitt, Schmitt & Clapman, 2001; Schmitt & Schmitt, 2012; Schmitt, 2010; Trembley, Baayen, Derwing & Libben, 2008). In Nation's most recent study (2006) the frequency-based approach is highlighted, yet this notion has tended to contradict his established view on vocabulary, highlighted in Nation (2001: 1-12, 2011: 12-13). Nation breaks vocabulary into four categories: high frequency words, academic words, technical words and low frequency words. Even so in this sense high frequency vocabulary is considered as an essential band for language learners.

Since vocabulary frequency in corpora is a continuum, breaking it into bands is more or less arbitrary. So, while Nation has two frequency categories, the first at the top of the vocabulary frequency scale and the low frequency for the rest, these have been given similar treatment. But it is necessary to also highlight the importance of the middle frequency range of the vocabulary. Schmitt and Schmitt (2012) illustrate the importance of mid-frequency vocabulary as a necessary frequency band for pedagogical purposes. In

Schmitt and Schmitt (2012) the high-frequency English vocabulary includes the most frequent 3,000 word families[1], and the low frequency band has been lowered to the word families beyond 9,000. These redefined boundaries are differently labelled than those viewed traditionally by researchers and language practitioners (Nation, 1990; Schmitt, 2000). Specifically, Nation's (2000) Most Frequent Words list where high-frequency vocabulary consists of the 2,000 most frequent word families, and the low frequency vocabulary is ranged from any words beyond the 10,000 frequency level. The frequency boundaries by Schmitt and Schmitt (2012) are quite similar to Kuiper, Columbus & Schmitt's (2009) frequency of lemmatized verbs where the three frequency bands are structured in the same way but the frequency range is ranked into four categories, with high frequency vocabulary divided into light (or de-lexicalised) verbs (Grimshaw, 1990) and non-light high frequency verbs. The four categories are classified as shown in Table 2.1

Table 2.1 *Categorization of verbs in Kuiper, Columbus & Schmitt (2009)*

Category	**Frequency criterion**
High frequency light verbs (HL)	Appearing in the top 1-3,000 words in the MFW lists (as words). Note that light verbs are also higher in frequency than the other high frequency verbs
High frequency lexical verbs (H)	Appearing in the top 1-3,000 words in the MFW lists (as words)
Medium frequency lexical verbs (M)	Appearing in the 3,000-5,000 word list in the MFW lists (as words)
Low frequency lexical verbs	Not appearing in any lists

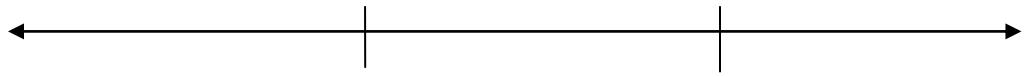
[1] A word family includes a root form (select), its inflections (selected, selecting, selects) and its derivatives (selection, selective, selectively, preselect).

A study by Cobb (2007) has provided significant insight supporting the frequency distribution. This study, which is related to frequency and incidental acquisition, targets about 30 words where the distributions are 10 from each of the 1,000, 2,000 and 3,000 levels. The aim is to see how often they occur in a 517,000 word extract of the Brown written English corpus. The results suggest that the 3,000 level (rather than 5,000) is the lowest frequency which can be considered to be in the high frequency level in terms of word learning. The second section of the study utilizes a corpus of just under 300,000 words which is assembled from Jack London's stories. The results reveal that the stories contain 817 word families at the 3,000 level, yet only 469 of them occurred six times or more while the remaining word families occurred five times or even fewer. It further illustrates that 3,000 word level is the optimum level where high frequency is to be ranked. In fact Schmitt & Schmitt (2012) observe that 3,000 word families represent an important milestone in language development, and thus identify the 3,000 level as a significant criterion for pedagogical use.

The next word level, labelled as mid-frequency vocabulary (Schmitt & Schmitt, 2012), represents the words from 3,000 until 9,000. They further discuss that this word level is not often addressed pedagogically though it is seen to have considerable importance and benefits for pedagogical purposes. This level is also addressed by Coxhead (2000, 2011) as the Academic Word List (AWL).This AWL is extensively used in English for academic purposes (EAP) classrooms, in various vocabulary tests (Schmitt, Schmitt, & Clapham, 2001; Nation, 2001; Coxhead, 2006; Schmitt, 2010) as well as being used as one of the major resources for researchers. This list consists of 570 word families which

are extracted from a written academic corpus of 3.5 million running words. To make this clearer, the graph below represents a common conceptualization of vocabulary frequency (in Schmitt & Schmitt, 2012)

Figure 2.1 A common conceptualization of vocabulary frequency

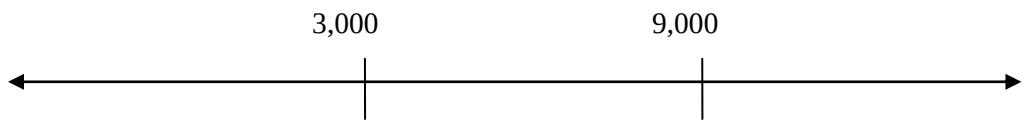


High frequency vocabulary (frequent in all discourse)	AWL (frequent in academic discourse)	Low frequency vocabulary (rare in all discourse)

Note:
AWL- Academic Word List

Based on this conceptualization, Coxhead's Academic word list is ranked as 'middle' frequency of the vocabulary and the level subsequent to that band is low frequency vocabulary. This principle is 'parallel' to Schmitt & Schmitt's (2012) approach to vocabulary frequency where there are three vocabulary levels and the nature of the frequency distribution is as represented below:

Figure 2.2 Schmitt & Schmitt's (2012) approach to vocabulary frequency



High frequency vocabulary	Mid-frequency vocabulary	Low frequency

Yet, it cannot simply be concluded that low frequency vocabulary is the subsequent band beyond those two levels (high frequency and the AWL vocabulary), as the first 3,000 word families of high-frequency vocabulary also include members of the AWL (Schmitt & Schmitt, 2012; Hancioglu, Neufeld & Eldridge, 2008). The most recent and relevant studies by Schmitt & Schmitt (2012) and Nation (2006) have proven that the low frequency level can plausibly be ranked as vocabulary beyond the 9,000 band (9,000+).

The above literature supports the approach taken in the present study of making the relevant word frequency list by ranking the verbs in 3 frequency levels: high-frequency, medium-frequency and low-frequency. So, the present study will adopt the above frequency levels with the additional insight of splitting the higher end of the high-frequency level as 'high light' frequency verbs (Grimshaw, 1990). In addition to that, the present research adopted not the word families as the frequency distribution, but instead used lemmas.

2.7 Vocabulary acquisition and the effects of exposure

It is reported that the size of English language vocabulary is ranged from a half million to over 2 million words (Crystal, 1987), and the estimation of word family unit counting based on Goulden, Nation and Read's (1990) calculation in Webster's Third New International Dictionary (1963) found that it contained about 54,000 word families. Both are considered huge numbers either counted by individual units or by word families. The task of native acquisition of a usable proportion of such a large vocabulary can thus be seen as considerable.

The problem is greater for second language learners. Pavlenko (1996) claims that exposure to the culture and the L2 language may affect learners' language processing. It is said that the shared concepts and the cultural values of the L1 and L2 languages may influence and enhance the processing to be faster and more accurate (Kosmos, 2006; Masoomeh et al., 2012). In the case of L1 acquisition, difficult tasks or high-pressure but routine oral performances such as performing in real-time in language production, i.e. in the case of oral poetry, storytelling, auctions and sports commentaries, are possible and can be achieved through frequent exposure to the relevant varieties. Once the culture and traditions are assimilated these oral productions are able to be presented fluently. It seems therefore that frequent exposure enables speakers to produce quality language production. It should be the same case with frequent exposure to learning a second language, where frequent exposure may be the result of quality learning. According to Fitzpatrick (2012) exposure seems to elevate the progress of vocabulary acquisition. That study reveals how the experience of studying abroad impacts the sophistication of word knowledge and lexical organisation.

According to Schmitt (2000) vocabulary is acquired through two main processes known as *explicit learning* and *incidental learning.* Explicit learning focuses attention directly on the information to be learned, i.e. the study of words. This approach allows for more opportunities for acquisition to occur though it might be time-consuming. Some researchers tend to agree that explicit learning of vocabulary may be responsible for most L2 vocabulary learning (Laufer, 1991, 2001; Webb, 2008; Boers, Demecheleer, and

Eyckmans, 2004; Boers, Lindstromberg and Eyckmans, 2014; Lindstromberg and Boers, 2008a; Webb and Kagimoto, 2011). Yet, the later approach may centre the learning through exposure, where learners' attention is focused on the use of language. This type of learning can be achieved when learners use language for communicative purposes. Yet, this is considered a slower process and requires a learner to be exposed to a great number of texts as well as interactions in order to come across lower frequency lexical items. The lower frequency lexical items may also be genre related, i.e. phrasal lexicon of aircraft maintenance engineers (Newsome, 2005) thus it may not predict exposure well.

Incidental learning is a main approach of acquiring vocabulary in an L1 (Nation, 2001; Schmitt, 2000). However, for second language learners both explicit and implicit types are required (Nation, 2001; Schmitt, 2000; Waring and Takaki, 2003). The focus should be laid on teaching or learning a certain target of important words, i.e. the most frequent words in a language; technical phrasal lexical items should come later (Schmitt, 2000). Hence, the infrequent ones should be left to incidental learning, i.e. exposure. Indirectly, in order to accelerate incidental learning it is necessary to increase the amount of exposure to the target language and the required genres. It is seen that the effects of repeated exposure tend to offer gains of incidental vocabulary learning. However, exposure is commonly seen as a major problem in second language learning as insufficient exposure may affect vocabulary learning.

So, how much exposure do learners have to a particular item of phrasal vocabulary and how long should the exposure be to be sufficient for learning to take place?

There are various suggestions on how many encounters are needed to learn an unknown L2 word. Rott (1999) suggested that six encounters are needed where Horst, Cobb and Meara (1998) found that eight encounters are sufficient. On the similar consequence, Saragi, Nation and Meister (1978) suggested 10 encounters, Webb (2007) recommended that more than 10 encounters are required, and Waring and Takaki (2003) found that at least 20 encounters are needed to incidental learning of a word. In a recent study by Webb, Newton and Chang (2013) on incidental learning of collocation, it is found that the more often a collocation is encountered the better the chances of it being recalled in the post-tests. The results suggest that the number of encounters has a positive effect on learning, thus 15 times is considered as an ideal number of encounters. Durrant and Schmitt's study (2010) indicates that repetition may have an effect on learning collocations incidentally in the similar way that it does for single-word item.

It is found that context may have been the reason for the contrasting results of encounters needed to learn an unknown L2 words or phrasal items. And for sure a clear definition of context in research is required when comparing the results from different studies.

2.8 Vocabulary acquisition in Malaysia

In this section we look at Malaysian English in the context of various kinds of English. It is important to do so since it allows for an assessment to be made of what kinds of collocations we might expect speakers of Malaysian English to have acquired.

Kachru (1992) has categorized the spread of English into two different diasporas. The first category involves extensive numbers of migrations of speakers of English from British Isles to countries like Australia, New Zealand and North America, while the second category mainly involves Asian and Africa contexts, where the new sociocultural contexts were present and influenced the way in which English was spoken. Kachru has categorized the users of English based on a model named Concentric Circles (Kachru, 1985, 1992).

In this model there are three circles or groups labeled by Kachru: Inner Circle, Outer Circle and Expanding Circle. Countries like New Zealand, Australia, the UK and the US are grouped as the Inner Circle. Countries such as Malaysia, Singapore, India and Nigeria are part of the Outer Circle where English is used as a second language. Countries like China, Korea, Japan and Taiwan are grouped as the Expanding Circle because English is used among several other languages of communication. Thus, such a view has serious implications for the way English is conceptualized in the world context. As Kachru (1996) highlighted in his argument, what is now pertinent is 'Englishes' rather than 'English'. English has a multiplicity of norms, both endocentric and exocentric, with multiple identities in creativity, and separate sociolinguistic histories and contextual functions.

It is important to note that in outer circle languages, contact conditions have partially merged languages, cultures and values. While it is interesting to see that these contemporary New Englishes have much in common, they also have become unique and

varied in terms of grammatical innovations and tolerances, lexis, pronunciations, idioms and discourse (Kachru, 1992). In fact, there are collocations with different frequencies of occurrence. As an evidence, a list of collocations (refer Table 3.1) with different frequencies listed on page 63 of this thesis. That list suggests that even the collocations are the same as standard English in form, they have different frequencies of occurrences.

In this context the present study will examine the use of restricted Verb-Noun collocations, such as *does wonders, crack (a) joke* and *give (a) hoot* by second language users of one of these new Englishes. The study focuses on the factor of exposure to English, both the new English of Malaysia and an inner circle English, to see whether the length of exposure to English affects the acquisition of restricted collocations. This is because knowing a language involves more than having the knowledge of its central linguistic elements. It also incorporates the knowledge of the idiosyncratic properties of a language, such as formulaic sequences, which often reflect the culture of its native speakers' way of thinking, feeling and interacting (Makarova, 2010). In this sense, the researcher's aim is to deepen the understanding of Malaysian speakers' use of English collocations with the emphasis on the second language users' written language.

It is quite common to observe in most English proficiency subjects at tertiary level in Malaysia that learning collocations is part of the syllabus. As for example English subjects like English for Communication 1and English for Communication 2 in University Utara also seem to take collocational learning seriously. Browsing through primary school's text book (in this case is Year 3 English textbook) verb phrase might not be a common exercise but at least supplying missing verb activities are found as quite

normal to be practiced. What is quite common is that the use of cloze testing is part of the most commonly testing procedures. Thus, it is worth investigating these areas as they have received considerable attention by this group of second language users.

2.8.1 Language policy and situation in Malaysia

The issue of language and nation building

This section provides background information on language policy and the situation in Malaysia. Malaysia is a multicultural country and is used to experiencing the problems of standardization and standard language which include spellings and rules of speaking (Asmah Haji Omar, 1992). Added to that, the problems have created a beautiful linguistic scenery in Malaysia.

In order to understand the acquisition of a second language such as Malaysian English, it is important to see such acquisition in its socio-cultural context. The language situation in Malaysia may be described as multilingual and broadly diglossic (Asmah Haji Omar, 1992), as the population of Malaysia comprises different ethnic groups, namely the Malays, Chinese, Indians, others (Portuguese, Baba's etc.) and the Indigenous peoples. The verbal and speech repertoires of these different ethnic groups are varied and would include a native tongue and a second or third language, besides dialects. For this diverse linguistic community, Malay is the sole national and official language (Hafriza Burhanudeen, 2006). Malay has been legislated as the official and national language of Malaysia and should be used by individuals, by groups, private or public agencies in every field and activity of life. Ooi (2000) claims that 'Malay is both the de jure and de facto national language in Malaysia' (pp.196).

English played the role of official language in Malaysia until ten years after the Malayan Independence of 1957 (Asmah Haji Omar, 1992). The Language Act of 1967 established English as the second most important language for Malaysia and it is currently taught in schools. The Malaysian government is aware of the fact that English is the global lingua franca, or language of wider communication (LWC) and therefore it is acknowledged as the language used in business, science and technology. English, French and other languages of wider communication (LWCs) are considered the world's business languages and provide effective means of communication.

Although English came to be used in Malaysia with colonial occupation, the end of the colonial period did not mean that English disappeared. The government itself was responsible for allowing the continuity of English in Malaysia for certain official events and in certain domains. The fact that English has been retained shows the intention to continue its use for the purpose of nationism.

The concept of nationism and nationalism in language planning as well as language use has been raised by Fishman (1968). Nationalism has been defined as 'the process of transformation from fragmentary and tradition-bound ethnicity to unifying and ideologised nationality' (Fishman, 1968: 41). Asmah Haji Omar(1992) points out that nationalism is the matter where self-identity and group identity are developed through a common language, while nationism is a matter of efficiency or group cohesion. In this sense, Asmah believes that Malaysia has achieved nationalism through its national language, Bahasa Malaysia or Malay, and at the same time English has made nationism possible. What's more, nationalism and nationism should not be treated as two distinctive

concepts, rather they are supposed to embrace one another. However, when language is of concern the concept of nationalism and nationism can become a matter of conflict (Alis Puteh, 2010). The issue of symbolic and pragmatic purposes will arise in implementing the national language of a country. Apart from being a global language of communication, English at the same time not does not want to be seen as downplaying Malay as the national language (Shameem Rafik-Galea and Mohd Salleh, 2003).

Malaysian Education policy

In the Malaysian context, Bahasa Melayu or Malay has affirmed its role as the national language as well as being the medium of instruction for schools and other public education institutions. Yet, the reality is that the private education institutions use English as the medium of instruction. This might be due to the fact that English is still seen to be the language necessary in order to compete in this age of globalization. However, to place English as a second language is considered by some scholars as a major mistake. As Shameem Rafik-Galea and Mohd Salleh (2003) suggest that the relegation of English to that of a second language caused the standard and use of English to decline drastically. Thus, Malaysia like many countries belonging to the Outer Circle has to cope with the declining standards of English.

Such views have led to the realization of the need for a change in language planning and policies in order to face globalization. As Malay is given the liberty to grow without restriction, English has also been promoted as a co- language which contributes to the nation's growth and success. Shameem Rafik-Galea and Mohd Salleh (2003) suggest that this situation is thus leading to promoting Malay-English knowing-bilinguals. In fact, in

the year 2003 the Ministry of Education gave English the privilege of becoming pre-eminent in some domains. The teaching and learning of science and mathematics is in English (or PPSMI to use the Malay acronym), was introduced by the fourth Malaysian Prime Minister, Tun Dr Mahathir Mohamad. The policy aimed to improve the command of English as well as to prepare the nation for the rapid changes of the future. Initially, the implementation only involved year one of both primary and secondary school levels, and lower six form of upper secondary level. It was expected that the cycle would be completed by the year 2008, and by the year 2009 and 2010 all public institutions at tertiary level would continue using English for these two subjects. Because the sources of science and mathematics are in English, it was assumed that the students would at the same time master the language. However, this policy had been 'rejected' by most Malaysians as Malay should be taught and learnt at schools and tertiary levels. In 2011, the Ministry of Education announced that Malay was officially the language for teaching and learning mathematics and science, and it is going to be fully implemented by 2016 (press conference, 9 November, 2011). This seesawing language policy seems to have no end.

This study is indirectly looking at the effects of PPSMI policy on the learners of English in Malaysia. Since it was implemented in the year 2003, a total of ten years of education, it is worth looking at. The form five students who are at the age of 17 in this study have experienced the policy from year 1 of their primary school level. This study may not be exclusively looking at the effects of using English for science and mathematics but since this study investigates exposure to English, these subjects have obtained ten years of

learning English at schools. On the strength of this exposure it may be supposed that more restricted collocations have been acquired.

2.9 Theoretical framework of the study

This section discusses the nature of collocation and restricted collocations as well as looking at the theories related to the study.

This study adopts Sinclair's (1991) model of the way words occur in a text. As mentioned before, Sinclair has outlined the distinction between the *open-choice principle* and *the idiom principle.* The *open-choice principle* is where language text is seen as a series of choices where the only limitation on choice is grammaticalness (pp. 109). This principle is often referred as the 'slot-and-filler' model with the idea that language is creative and operates simultaneously on several levels. Therefore, a wide variety of possible words can be filled into each 'slot'. Sinclair claims that this could probably be traditional way of describing language. The *idiom principle* proposes that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (p. 110*).* The idiom principle illustrates the fact that there are patterns or regularities in how words co-occur with each other. Within this view, recalling the earlier discussion, collocation is defined as the occurrence of two or more words within a short space of each other in a text (pp. 170). The pervasive nature of the idiom principle is significant enough to highlight the importance of collocation. This also suggests that there may even be a larger number of phrasal items as compared to individual words, as the idiom principle is argued to be dominant over the open-choice principle.

Apart from Sinclair, Nation (2001) defines collocations as closely structured groups whose parts frequently or uniquely occur together. And collocations are expected to contain some element of grammatical or lexical unpredictability or inflexibility (p. 324). For this study, this research applies the term 'collocation' in the Firthian sense of the habitual co-occurrences of a group of words (Sinclair, 1991; Ooi, 1998).

Together with the above concept, Howarth's model of a continuum (1998) will be applied. Howarth (1998) assumes that lexical items in these categories do not have definite boundaries, rather they are ranged on a cline from pure idioms to free combinations as outlined in Figure 2.1.

Figure 2.3 Howarth's model of continuum (Howarth 1998: 28)

Pure idioms	Figurative idioms	Restricted collocations	Free combinations
[blow the gaff]	[blow your own trumpet]	[blow a fuse]	[blow a trumpet]
[under the weather]	[under the microscope]	[under attack]	[under the table]

The leftmost extreme of the continuum contains pure idioms, followed by figurative idioms and restricted collocations. The rightmost extreme contains free combinations also known as pure syntactic constructions. These categories do not have definite boundaries, but rather reflect shifting proportions of semantic and syntactic characteristics. These categories are later used as guidelines in selecting the list of restricted collocations for the testing procedure in Study 1. The selection is later done manually (concordance search

within the corpus i.e. NST corpus) though there are several statistical ways to measure association e.g. mutual information. MI is a statistical measure proposed by Church and Hanks (1990) to estimate the degree of association between words. Unfortunately, the present study appears not to use it since the manual search done for the study is assumed to be sufficient and accurate. In other words, the approach for the collocational selection list is based on the phraseological standpoint where it uses lexical criteria such as the degree of fixedness.

Lexical items can be broken into groups based on their compositionality- whether or not the meaning they express is the sum of the meaning of their parts. Non-compositional items are idioms, while among the compositional items are collocations and clichés. The meaning of a collocation can be predicted (semantically compositional), however it is nevertheless particularized (Mitchell, 1971). In contrast, idioms are different in that they have meanings that cannot be predicted from the meanings of the parts. The term idiom and idiomaticity carry a rather different meaning. Yet, when we talk about idiomaticity, it is actually a matter of degree.

Apart from Howarth's model of continuum, this study is adapting the notion that most researchers agree upon, namely that there are two basic kinds of collocations: grammatical or syntactic collocations, and lexical or semantic collocations (Bahns, 1993; Biskup, 1992; Benson, 1985). Benson's fundamental distinction on the basis of word classes will be used in classifying collocations. Benson (1985) has classified collocations into two major components; grammatical collocations and lexical collocations.

Grammatical collocations consist of a dominant word, usually a noun, an adjective, or a verb, plus a dependent word such as a preposition or a grammatical structure such as an infinitive or clause. Lexical collocations, in contrast, consist of content words only, such as nouns, verbs, adjectives, and adverbs not including prepositions, infinitives, and clauses. Lexical collocations consist of two 'equal' components, such as Verb + Noun, or Adjective + Noun.

The present study focuses on the restricted collocations that are headed by verbs, not auxiliary verbs. The researcher will only consider Verb + Noun collocations in verb plus complement constructions as part of this study. Table 2.2 and Table 2.3 illustrate the two major components of collocations.

Table 2.2 The subtypes of grammatical collocations (Benson, 1985)

Type of Grammatical Collocations	Examples
Verb + Noun	*(to) make a decision (to) take notice*
Adjective + Preposition	*different from, curious about*
Adjective + Preposition + Preposition	*fed up with*
Preposition + Noun	*for sale, on time*
Dative movement transformation	*She sent the book to him/ She sent him the book*

Table 2.3 The Subtypes of Lexical Collocations. (Benson, 1985)

Types of Lexical Collocations	Examples

Verb + Noun (pronoun, prepositional phrase)	*(to) reach a verdict* *(to) revoke a license*
Adjective + Noun	*reckless abandon*
Noun + Verb	*alarms go off*
Noun + of + Noun	*a bunch of flowers*
Adverb + Adjective	*deeply religious*
Verb + Adverb	*(to) apologize humbly*

As already mentioned in Section 2.2, verb+noun collocations are targeted because they are the most frequently used combinations (Bahn & Eldaw, 1993; Bahns, 1993; Biskup, 1992). This fact suggests that more attention be placed on these combinations as there exist struggles among second and foreign language learners.

Superlemma theory

This study adopts the model of idiom representation in speech production, namely the *superlemma model* (Sprenger et al., 2006; Kuiper et al., 2007). This model merges the model of idiom production (Levelt, 1989; Levelt & Meyer, 2000) into a contemporary model of language lexical access. Levelt (1989) proposes that the mental lexicon is the organization of lexical knowledge in the mind and it allows access to various types of linguistic information at different stages of speech production process. Levelt further asserts that the mental lexicon consists of interconnected nodes that encode lexical information at various level of abstraction. Superlemma theory assumes that a phrasal lexical item has a single lexical concept. The superlemma is the unitary representation of

a lexical phrase which consists of constituent lemmas of the idiom and their unique syntactic properties. In other words, formulaic sequences are stored holistically. So what happens during speech production is that when a single lexical concept is activated, then its superlemma node is activated. The activation of the superlemma node in turn activates the lemma nodes of all its constituent words. So, once the superlemma is sufficiently activated, a user or learner may be able to retrieve the missing word to fill up the gaps in cloze test. This process provides evidence that cloze testing is an excellent method for investigating the acquired phrasal lexical items. In other words, if a language user is able to provide the missing word, he or she is assumed to have the knowledge of that particular expression. The process of retrieving the missing words involves a move from perception to production since it requires a user to fill up the slot with a selected word.

2.10 Chapter summary

This chapter has provided background on diverse definitions outlined to describe lexical phrases. Regardless of the various terms and names given, the major attention is on the same aspect of linguistic production; phrasal lexical items. Kuiper's comprehensive approach is best to describe the current research scope. For the purpose of the study, restricted collocations are opted to be explored as they play great importance in vocabulary acquisition.

It is also an interesting phenomenon to be observed as English came to be used in Malaysia with colonial occupation. This study provides an update of usage and supplies information on the current status of English in one of the New Englishes, as well as

looking at the effects of frequency and exposure on vocabulary acquisition of the standard English.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents details of the corpus procedures used in the two studies reported later in this work. It provides the overall process of obtaining the frequency of the verbs used in the cloze test for the study reported in chapter 4. It illustrates significant steps of the research as well as reflecting on the development of a corpus as one of the tools of the present study. There are improvements still to be made to the corpus since building a corpus is a complex process. The corpus quest details will be discussed in a subchapter of the thesis i.e. Section 3.1.

Since this study focuses on the acquisition of restricted collocations in Malaysian English, and no significantly large corpus is available to supply the frequency list for the cloze testing items, a new corpus of Malaysian English was required. With regard to the availability of English corpora in Malaysia, there are a few existing corpora developed by different academic institutions in Malaysia. Among those corpora are MACLE, COMEL, CALES and EMAS. The Malaysian Corpus of Learner English (MACLE) and the Corpus of Malaysian English (COMEL) (Knowles & Zuraidah, 2004) are developed by University Malaya. MACLE is a written corpus which consists of students essays, whereas COMEL is a spoken corpus which is still under development. The Corpus Archive of Learner English Sabah-Sarawak (CALES) is an ongoing project which was started in 2003. The current number of words is approximately 400, 000, consisting of students' argumentative essays (Botley & Dillah, 2007). The English of Malaysian

Schools Students (EMAS) corpus currently has 472, 652 words of both written and spoken texts. The EMAS corpus was pioneered by a group of researchers at the Universiti Putra Malaysia (Arshad Abd. Samad et al., 2002). It sampled students' writing and speech in primary and secondary schools all over Peninsular Malaysia. Given that the existing English corpora in Malaysia are still under development, and that the current number of words or the size is not large enough for this study, a newly designed corpus was required.

The process of building the present corpus involved a number of stages (refer Figure 3.17). The written texts of the corpus are press news reports and press editorials. For this purpose, the English language newspaper, *News Straits Times*, was chosen to represent other local English newspapers, and is regarded as representing formal written Malaysian English. According to Schmitt (2000) newspapers are often a convenient source of authentic L2 materials, especially when English is the target language. The corpus contains 19 million words. Though News Straits Times is not the only exposure Malaysia students have to Malaysian English yet the News Straits Times distribution of collocations is considered a good proxy as NST corpus is the only large current corpus of Malaysian English. In fact it is the only available source of frequency data. The main available source of Malaysian English input for Malaysian students is probably spoken vernacular Malaysian English for which there is no corpus.

To generate the frequency data, the *Wordsmiths Tools 5*, a word search programme, was employed to do the search for phrasal lexical items, specifically restricted collocations, which this study used as data.

The next stage of the project was developing the cloze tests items. The test focused on the influence of both the learner's exposure to English and the frequency of lexical items, as a proxy measure of exposure, on the acquisition of restricted collocations. In this case, where learners' Malaysian English is studied, the frequency list of verbs gathered from the NST corpus contributed to constructing the cloze test. The collocations used in both cloze tests (Studies 1 and 2) are common to standard English and Malaysian English. Those collocations are extracted from the NST and BNC corpora. The NST or *News Straits Times* is not in standard English as indicated by numerous loan words from Malay e.g. masjid (frequency-362), tudung (114), kari (86), melayu (589), kambing (92) and keris (92). And another fact is that the verb frequency of Malaysian English extracted from the NST corpus is not altogether standard English like *The Straits Times*, an English language daily broadsheet newspaper owned by Singapore Press Holdings. Apart from that, the frequencies of the verb phrase used for the Cloze test 1 which are extracted from NST corpus are different in PIE (Phrases in English). The following 5 examples (from the high light frequency band) in Table 3.1 indicate that the frequency of the same restricted collocations for Study 1 is different in BNC.

Table 3.1 The frequency of restricted collocations in BNC and NST corpus

Restricted collocations	Frequency of RCs in NST corpus	Frequency of RCs in BNC corpus
Does wonders	32	103
Make a fast buck	20	5
Taking a big risk	143	14
Get a grip of oneself	6	10
Give a hoot	3	15

As will be discussed in Chapter 4, the cloze test was administered to four different groups of students; two groups of students in Malaysia, one group of Malaysian students studying in New Zealand and one group of native speakers of New Zealand English. This cloze test was used for the first study which is described in Chapter 4 of the thesis. As for Studies 1 and 2 (Chapter 5), the test design used was a replication of that found in Kuiper, Columbus, & Schmitt (2009). Both cloze tests' head verb frequencies were selected based on four frequency bands of i.e. high frequency light verbs, high frequency verb, medium frequency verbs and low frequency verbs. Cloze test 1 was designed for Study 1used Malaysian English (as another variety of English) collocations which was selected from NST corpus, while for Study 2, the test set was replicated from Kuiper, Columbus & Schmitt, 2009 and taken as representing the standard English. Both testing question sets were using two different vernacular stories with the selected restricted collocations appropriately embedded in trying to be attractive and natural enough for respondents to engage with the stories. Participants were provided with a short

questionnaire in the form of very simple demographic information, e.g. age and years of studying English.

The following details provide comprehensive particulars of the study conducted for the purpose of this research. The particulars provide information on the measurement and procedures as well as a specific subchapter on the corpus procedures.

3.2 Background: Creating the NST corpus

Like other corpus projects employed to obtain word frequencies, the creation of the NST corpus falls into a number of phases of work including a corpus compilation phase, annotation or tagging, followed by the wordlist making procedure itself, and finally the verb frequency list. In order to understand the design principles of the NST corpus it will be useful to present here the main ideas of the project.

There are two main motivations for the corpus project.

First, there is currently no large corpus of Malaysian English which offers a satisfactory representation of the lexicon of standard Malaysian English. This presents a barrier to fully understanding Malaysian English as a means of communication by all who use it as their native, second, or even foreign language.

Second, the existing Malaysian English databases and corpora have serious disadvantages not just because of their small size but also because they do not form a balanced corpus of Malaysian English. As mentioned earlier, they are compiled for small language

projects and may consist of several students' essays and few excerpts of news reports, i.e. ICE corpus of Malaysian project. This corpus is still being developed and is untagged.

The following discussion reports the rationale, background and the procedures of the creation of the NST corpus.

3.2.1 Permissions

News Straits Times (NST) is the main English newspaper read by Malaysians. All the texts or news reports were obtained from the 6 years of the newspaper archive from the *News Straits Times* press. The news obtained was from the years 2005-2010. For this research purpose, the News Straits Times Press has granted permission for their news reports to be used free of charge.

Securing permission is a practical problem and this is a sensitive area of law. When the permission is secured this is considered as a safeguard against exploitation and piracy. This could be done easily if copyright holders understood precisely why their texts were desired (Sinclair, 1991: 15)

3.2.2 Spoken and written corpus

Language scholars and teachers agree that the spoken form of the language is considered as a better reference to the fundamental organization of the language than the written form. According to Sinclair (1991) 'there is no substitute for impromptu speech' (p. 16). But since this study is chiefly looking at the acquisition of written English, written

corpora are best to referred to. For this purpose two major corpora are taken as the main sources, i.e. the British National Corpus (BNC) for standard British English and a newly developed corpus, the NST corpus for standard written Malaysian English.

3.2.3 Corpus size

Sinclair (1991), a pioneer of corpus studies, proposes this definition of a corpus, 'A corpus is a collection of naturally-occurring language text, chosen to characterize a state of a variety of a language'. But this leaves open a number of significant concerns. The most common concerns in corpus-based research are the size and balance of a corpus. The size of the corpus is one of the main concerns to most researchers. Sinclair (1991: 18) outlines that 'a corpus should be as large as possible, and should keep on growing'. Hitherto, defining 'large' seems to be ambiguous and inconsistent. Geyken et al. (2004) argue that a corpus of 100 million words is not sufficiently large for a reliable study of collocations. Furthermore even very large corpora are prone to either bias or lack of balance. Sinclair (1991: 19) argues that:

In order to study the behavior of words in texts, we need to have available quite a large number of occurrences. Again the statistics are against us, since we classify the occurrences in terms of 'uses' or 'meanings' we shall find the same kind of imbalance again. One of the uses will typically be twice as common as all the others; several will occur once only, and that is not enough on which to base a descriptive statement. This is why a corpus needs to contain many millions of words' (p. 19).

In relation to the size of corpus, the corpus of Malaysian English created for this study, i.e. NST corpus is definitely small in size as compared to other English corpora such as British National Corpus (BNC) and The Corpus of Contemporary American English (COCA). Since the NST corpus is used to extract the list of collocations used for the cloze test there might be some issues raised against the corpus. The issue of suitability of this small-scale, newspaper-based corpus of Malaysian English will be questioned as it is too genre-specific and too small to be a reliable proxy of the frequency at which the Malaysian English learners have been exposed to these particular collocations. The collocations are not fully extracted mechanically (and statistically) from the NST corpus. The corpus is too small to be reliable for the frequency counts of restricted collocations since 5 instances per million words is regarded by many researchers including Moon (1998) as the lower threshold for such frequencies. In fact, many phrasal lexical items do not reach that threshold in larger corpora. But, it is big enough for the frequency counts of verbs.

3.2.4 Balanced corpus

Ooi (1998) believes that a corpus should represent and include a sampling of the range of genres, speech and writing styles across the speech community. Furthermore, newspaper corpora are regarded as representing the language of everyday reading of a particular speech community as they provide the central and typical collocations used in the linguistic context concerned. Yet, the corpus created for the current research, NST corpus, cannot be categorized as *balanced corpus* because typically only a *general corpus* is designed to be balanced, by containing texts from different genres and domains

of use including spoken and written. The *Survey of English Usage (SEU) Corpus* pioneered by Randolph Quirk in 1959 is one pre-electronic corpus which has samples of a wide range of genres with different degrees of formality including interviews, lectures, seminars, monologues etc. Kennedy (1998: 20) describes the SEU Corpus, which contains texts produced between 1953 and 1987, as an example of a general corpus, or sometimes referred to as a *core corpus.* Other examples of general or generalized corpora are the British National Corpus (BNC), American National Corpus (ANC) and the Corpus of Contemporary American English (COCA).

There are also corpora which are designed for specific research projects. These corpora are referred to as *specialized corpora* which contain texts of a certain type and aim to be representatives of the language of this type (Kennedy, 1998; Bennet, 2010. Examples of specialized corpora are the Canterbury Corpus which has been collected since 1994 and is part of the archive of New Zealand English (NZE) speech, and the Michigan Corpus of Academic Spoken English (MICASE). Apart from these there are several other types of corpora which are part of specialized corpora, i.e. training corpora, test corpora, dialect corpora, regional corpora, non-standard corpora and learners' corpora (Kennedy, 1998). Therefore, the NST corpus can also be categorized as a specialized corpus as it serves to represent one type of English variety, i.e. formal written Malaysian English. However, spoken language may normally be the standard use in most corpora. But to develop a spoken corpus requires a more complex and slower process as it needs to contain transcriptions of complex phonetics and prosody features which do not occur in writing. In summary, the best practice is to opt for a suitable corpus for a particular analysis

because ‘particular corpora tend to be suitable for particular types of analysis and some corpora are simply not suitable for certain types of research’ (Kennedy, 1998: 21).

Next, I will discuss the tagging of the corpus. The Constituent Likelihood Automatic Word-tagging System (CLAWS) (Garside, 1987; Leech, Garside, & Bryant, 1994) was used for that purpose.

3.3 Tagging and lemmatization

3.3.1 Tagging

In dealing with problems which might affect the counting of linguistics items, researchers or analysts would usually treat word forms which are inflected as lexemes or lemmas (Sinclair, 1991: 41-42; Kennedy, 1998: 207). An example given by Sinclair is the word *come*. Other related words i.e. *come, comes, coming* and *came* definitely appear to be related, and are grouped under what is termed a l*emma*. Kennedy outlines *lemmatization* as a process of classifying or categorizing together all the identical or related forms of a word under a common headword. This task can either be performed manually, i.e. lemma by lemma, or automatically by using some kind of automatic routine. However, manual lemmatization is impossible and quite impractical if a corpus is large. Thus, a reliable grammatical tagger is required in order to speed the process and simultaneously gain an accurate output.

A tagger is required to annotate automatically every word in the corpus with a label or ‘tag’ to appear with a grammatical word class to which it belongs in context. Then, the

tagger will indicate to the *lemmatizer* the part of speech of each word used. An example of a tagger is TAGGIT, which was used to tag the Brown Corpus. According to Vine (2011) word-class tags have played an important role in natural language applications in work related to speech recognition and information retrieval (p. 72). Tagging also plays a part in educational contexts, where it enables us to make use of frequency data in making decisions in language teaching and learning.

As mentioned above, before the lemma can be sorted, the texts need to be tagged. To start the process, all the news reports in the original text files of the *News Straits Times (NST)* needed to be changed to Plain Text before they could be tagged by a tagger like CLAWS (Garside, 1987; Leech, Garside, & Bryant, 1994). This tagger is an automatic part-of-speech (POS) tagger whose tagging principle is based on a probabilistic matrix. The LOB- Lancaster-Oslo/Bergen Corpus was tagged using CLAWS 1, which was the earlier version of this tagger, and BNC was tagged using CLAWS 4. The tagset which was used to tag the NST corpus was known as the C7 tagset. Several tags out of 132 tags assigned to the word forms in the NST corpus are shown in Table 3.1.

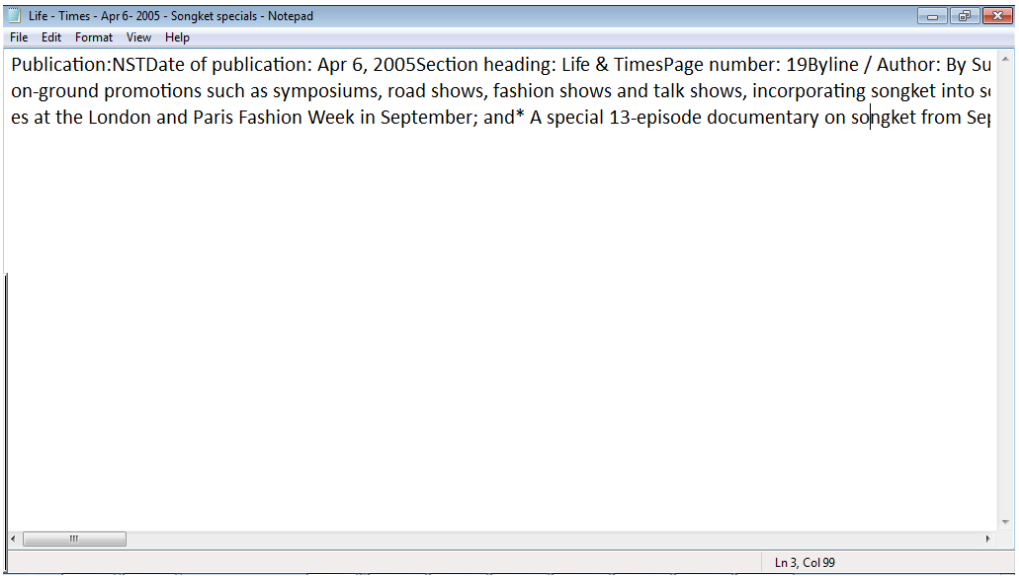
Table 3.2 The tagset used for tagging the NST corpus

VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	Were
VBDZ	Was
VBG	Being
VBI	be, infinitive (To be or not... It will be ..)
VBM	Am
VBN	Been

VBR	Are
VBZ	Is
VD0	do, base form (finite)
VDD	Did
VDG	Doing
VDI	do, infinitive (I may do... To do...)
VDN	Done
VDZ	Does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	Having
VHI	have, infinitive
VHN	had (past participle)
VHZ	Has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)

Later, the tagged texts can be processed or lemmatized using Wordsmith Tools 5.0. The diagrams below show some examples of news reports which are in Plain Text form. The following figure 3.1 shows the related examples.

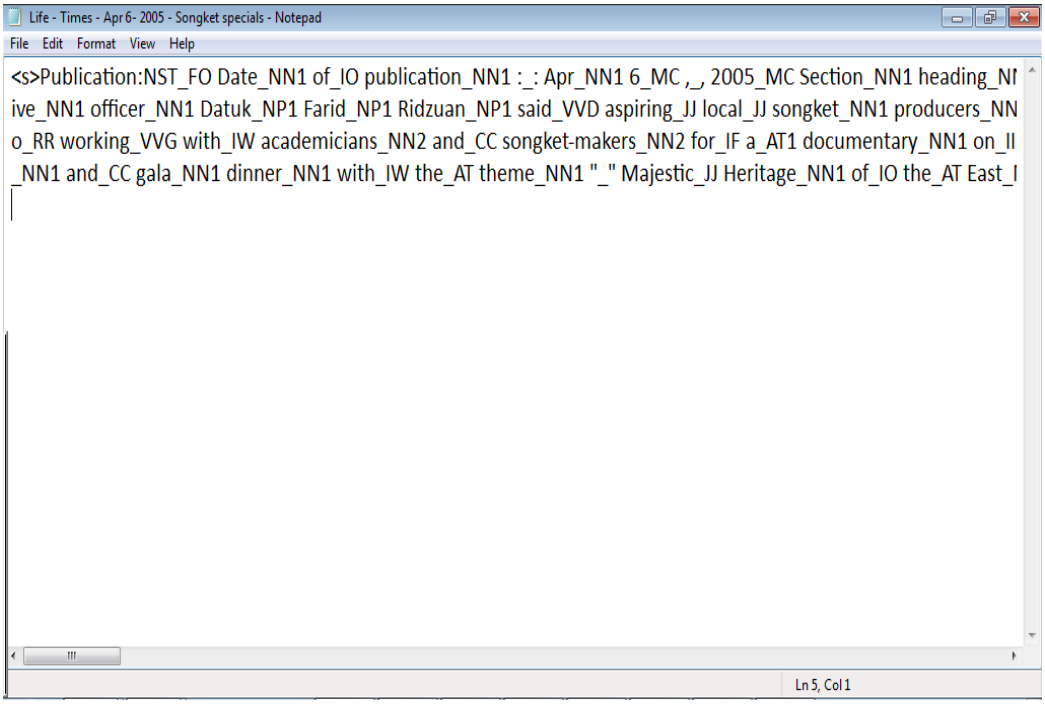
Figure 3.1 An example of news report from *NST* in Plain Text form



Automatic word-class tagging of corpora, which is also called Part-of-Speech tagging, or grammatical tagging, is the most common form of corpus annotation. In this case, it is the first form of annotation to be developed by UCREL at the University of Lancaster. It has been continuously developed since the early 1980s and has consistently delivered 96%-97% accuracy. However, the precise degree of accuracy may vary according to the type of text.

So, for this study, the corpus was tagged using CLAWS in order to be lemmatized for the purpose of obtaining word frequencies. The following table shows some examples of tagged text.

Figure 3.2 An example of a tagged text



3.3.2 Lemmatization

This might be too early to discuss the lemmatization process. Yet, lemmatization has been seen to cause a slight problem so it is worth mentioning at this stage. The tagging probably caused this problem to occur.

The lemmatization was complex as not all verbs were tagged accurately and some therefore had to be counted manually. In order to check accuracy, the original counting was also checked against frequency data from CELEX program. The same corpus was used, yet different software was utilized. With minor frequency count exceptions the frequency count was confirmed. For example the frequency of the verb ‘take’ has 33,991 occurrences, while ‘take’ in CELEX has 33,377.

CELEX was also used to 'filter' all other 'non-VERB' items in order to establish the verb frequency list only. Refer to Table 3.6 for details of the verbs frequency list created by CELEX. According to Vine (2011) automatic tagging may also not be accurate, i.e. comes with high rates of errors. Vine's study, which uses high frequency multifunctional words in the Wellington Corpora of Spoken and Written New Zealand English, has shown that the word-class tagging of the words was not as accurate as expected. She made a comparison between the unchecked automatic word class tagging and a manual analysis of random samples of uses of three high frequency multifunctional words, *as*, *like* and *so*. The results of the comparison showed rather high error rates in automatic word-class tagging of these sorts of words. This suggests that one should be cautious about interpreting word-class data derived from an automatic tagger.

However, before the process of lemmatization is further discussed I will briefly display how the wordlist was created using Wordsmith Tools 5.0. This is because the first attempt was to utilize this particular tool and only later CELEX was used for counterchecking of the frequency range. The following will briefly lay out the process of creating the wordlist.

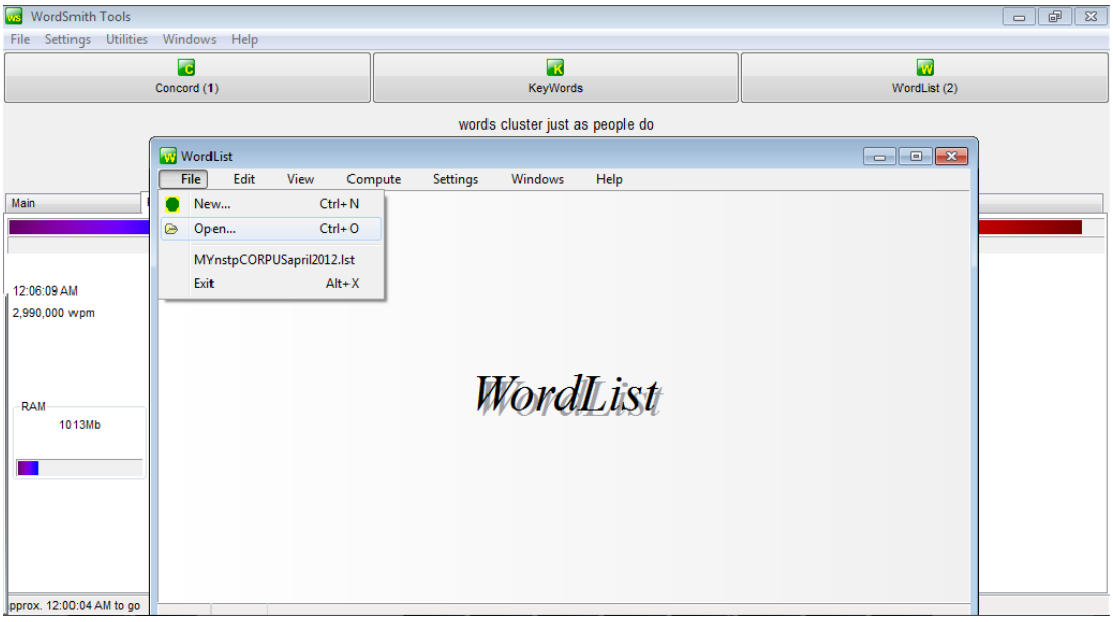
3.4 Creating a wordlist from the NST corpus

In order to obtain the frequency of the lexical verb forms in the NST corpus, a wordlist was established. As mentioned earlier, the task utilized Wordsmith Tools 5.0 with the *WORDLIST* making procedures, after which the verb only frequency could be counted.

The following is a brief procedure of obtaining the wordlist. The figures provided are meant to describe the procedures involved.

Step 1- Choose ***WORDLIST*** option menu (one of the options among the other two: ***CONCORD*** and ***KEYWORDS*).**

Figure 3.3 Step 1of NST wordlist procedure



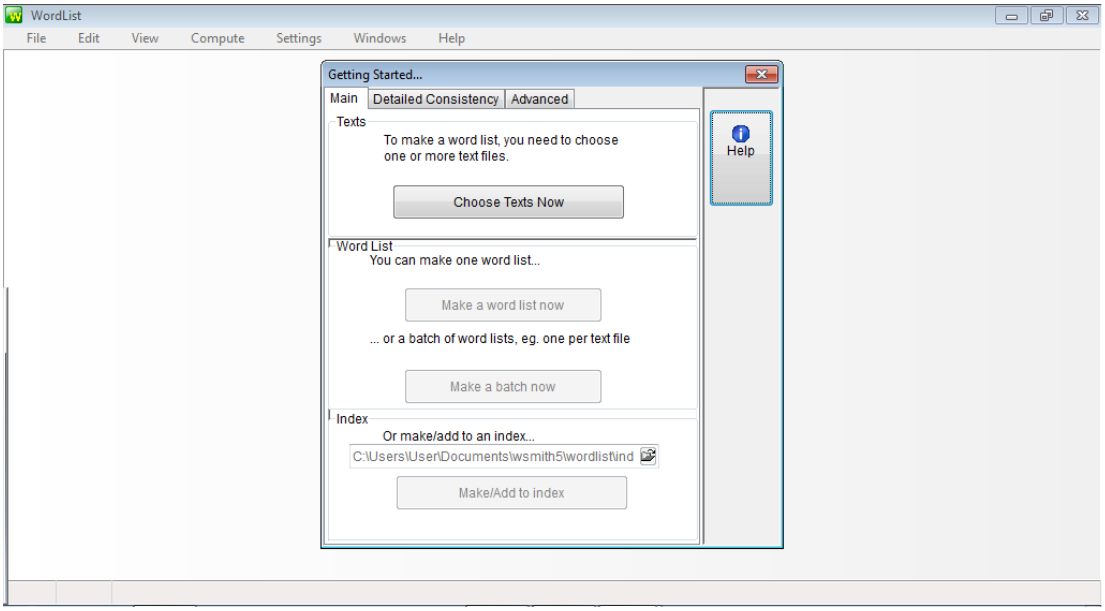
Step 2- There are 2 options for the next steps of the procedure, either

(i) Select existing files/corpus created, by opt for the ***OPEN*** button, or (above)

(ii) Opt for ***NEW*** button, and a new menu bar will pop up (below)

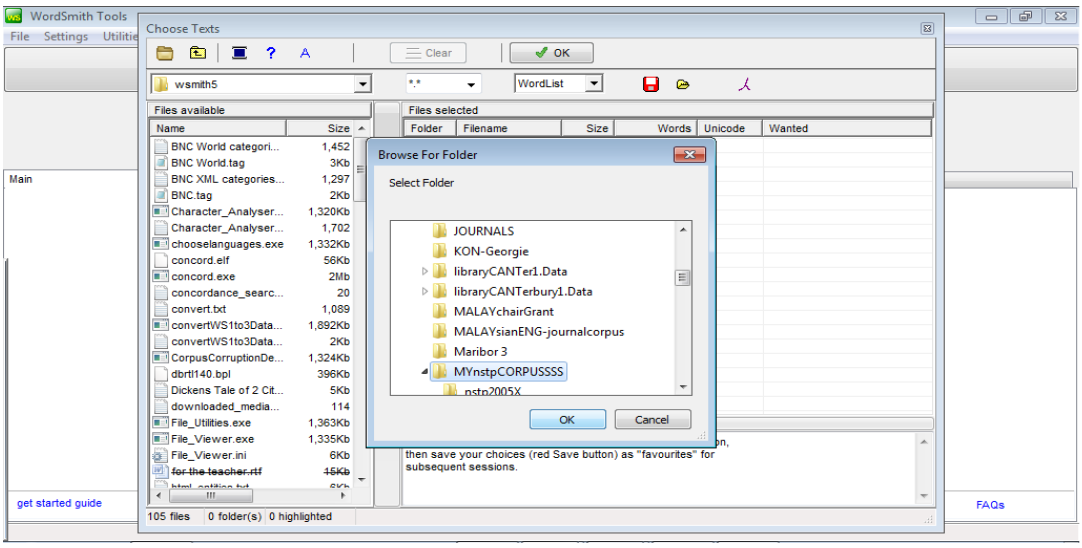
****GETTING STARTED -----MAIN-----CHOOSE TEXT NOW***

Figure 3.4 Step 2 of NST wordlist procedure



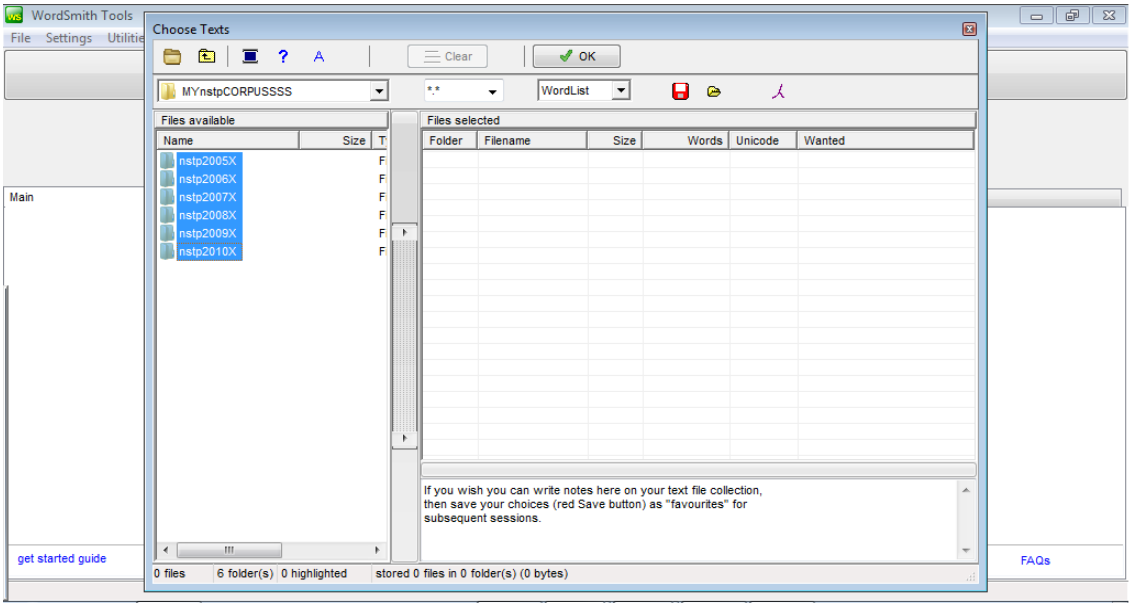
Step 3- Another menu bar will pop up-***CHOOSE TEXT*** (follow the steps below). A folder or corpus is selected from the drive, i.e. CoME-MYnstp corpus.

Figure 3.5 Step 3 of NST wordlist procedure



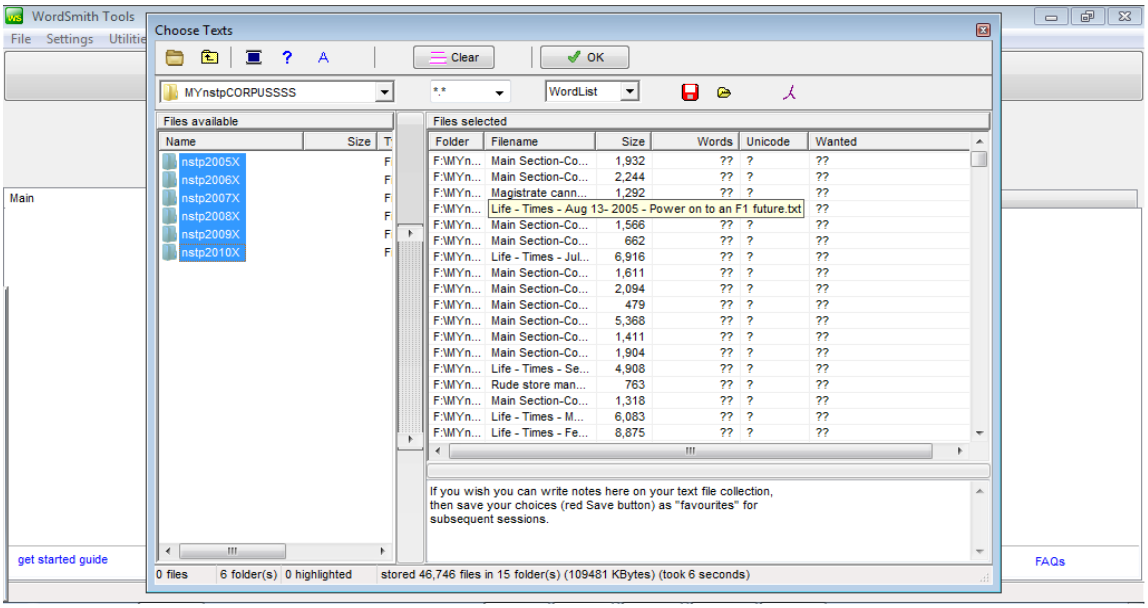
Step 4- Those files are marked as chosen, and later transferred to the right window. In other words, they are selected and copied to produce a frequency list.

Figure 3.6 Step 4 of NST wordlist procedure



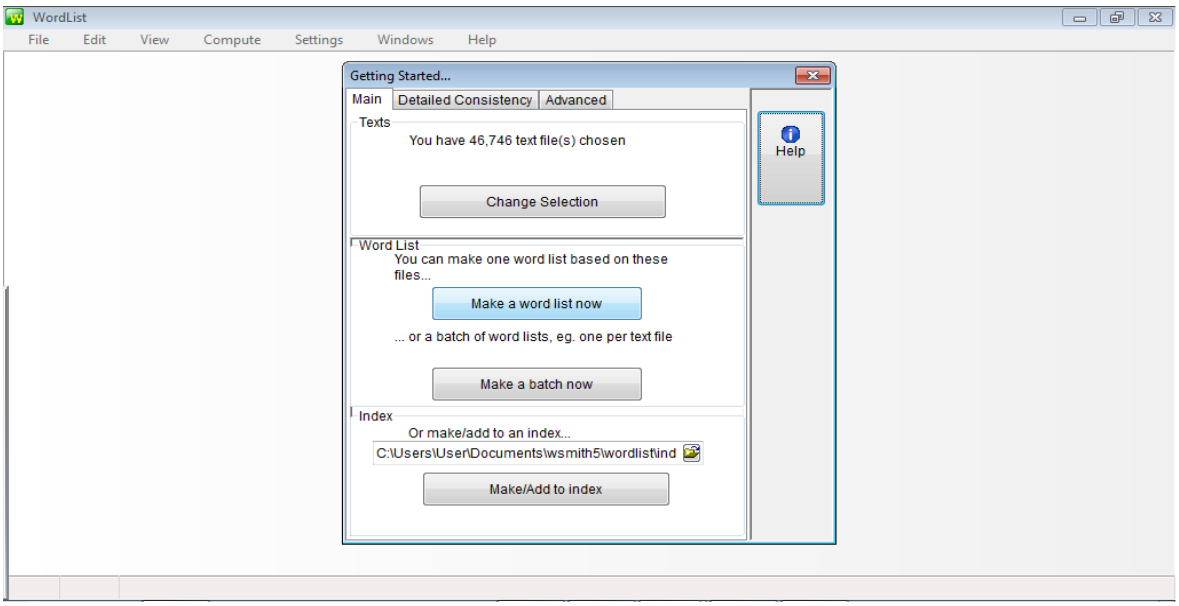
Step 5- This is the result after the files or the whole corpus had been copied and selected. They appear as individual files.

Figure 3.7 Step 5 of NST wordlist procedure



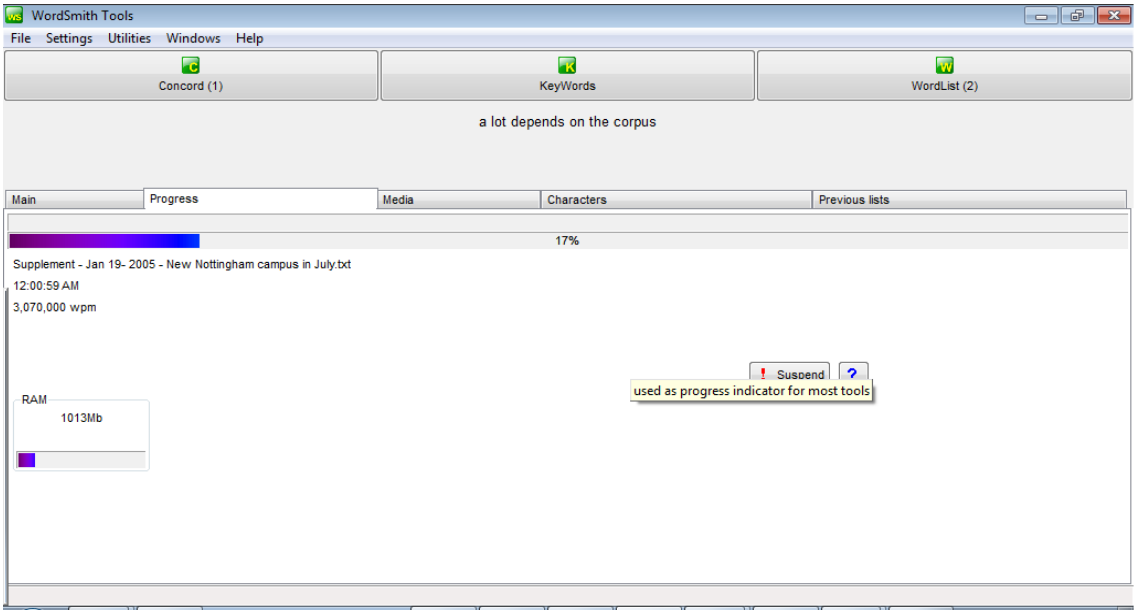
Step 6- After the button OK is pressed then this menu will appear. ***GETTING STARTED*** menu---***MAIN***---You have 46746 text file(s) chosen for the next procedure.

Figure 3.8 Step 6 of NST wordlist procedure



Step 7- After the button of ***MAKE A WORD LIST NOW*** is selected, you will see a progress bar on the main page. You can check the status of the task by checking the percentage(%) bar, i.e. 5%, 32% or 98%.

Figure 3.9 Step 7 of NST wordlist procedure



Step 8- Once the ***WORDLIST*** has been created, and try to Save the folder/corpus. In this case the corpus was saved under the name MYnstpCorpusApril2012 (in Document/ Libraries). The procedure yielded about 19 million token (running words in texts) and 18 million of the tokens were used for the wordlist. The following is the statistical detail:

Table 3.3 Statistics of NST corpus

Details	Statistics
File size	112 109 144
Token (running words in texts)	19 037 924
Token used for the wordlist	18 292 088
Types (distinct words)	154 549
Type/token ratio (TTR)	0.84
Standardised TTR	38.51
Standardised TTR std. dev	63.80
Standardised TTR basis	1000
Mean word length (in characters)	4.73
Word length std dev.	2.54
Sentences	46 746
Mean (in words)	391.31
Std dev.	1 773.70
Paragraph	46746
Mean (in words)	391.31
Std dev	1773.70

3.4.1 How the verb frequency count was established

Once the wordlist has been established, a general wordlist with frequency was generated. The next aim was to obtain the verb only frequency list. So, the lemmatized corpus in theory would be able to provide such a list. However, as mentioned earlier in section 3.3.2, the lemmatization process was complex and the corpus was potentially not tagged accurately. Therefore, the CELEX program was used to check against the frequency data obtained from Wordsmith Tools 5.0. So, when a list of verb frequencies was confirmed, it was later divided into three major verb bands; high frequency, medium frequency and low frequency.

After several selection stages, 20 head verbs were chosen. Howarth's model of continuum (1998) was used for the selection. These categories were later used as guidelines in selecting the list of restricted collocations for the testing procedure in Study 1. The selection was later done manually (concordance search within the corpus i.e. NST corpus) though there are several statistical ways to measure association e.g. mutual information. This study appears not to use it since the manual search done for the study is assumed to be sufficient and accurate. In other words, the approach for the collocational selection list is based on the phraseological standpoint where it uses lexical criteria such as the degree of fixedness.

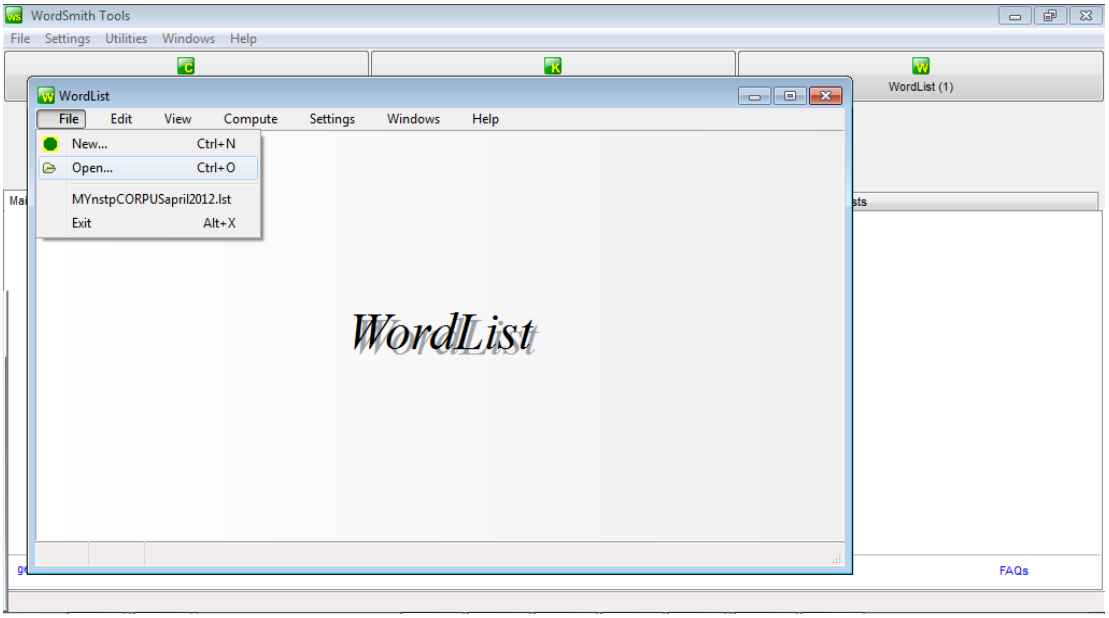
Based on the frequency counts of verbs, few head verbs were shortlisted e.g. take, do, give, make etc. Since it was too random to make selection, a frequency criterion (bands)

was established to make the selection easier e.g. high frequency light verbs and medium frequency verbs.

Based on these shortlisted head verbs, 20 strings were decided on and checked for frequency in NST corpus. The following is an example of how the frequency of a string or a restricted collocation was done using Wordsmith Tools 5.0.

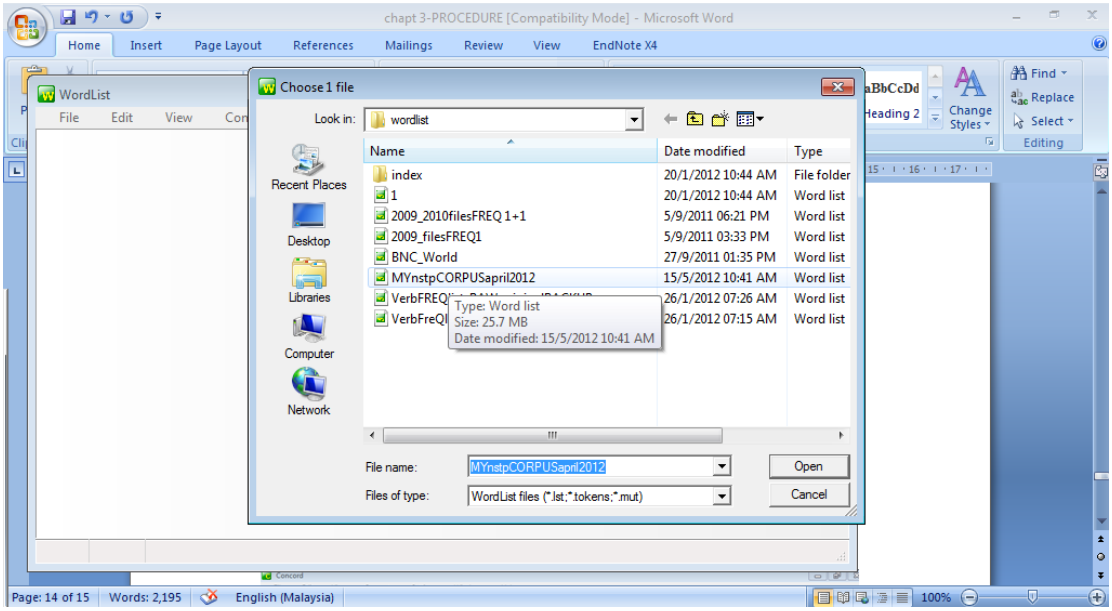
Step 1- The ***WORDLIST*** option or program is selected.

Figure 3.10 Step 1 of the restricted collocations frequency search in NST corpus



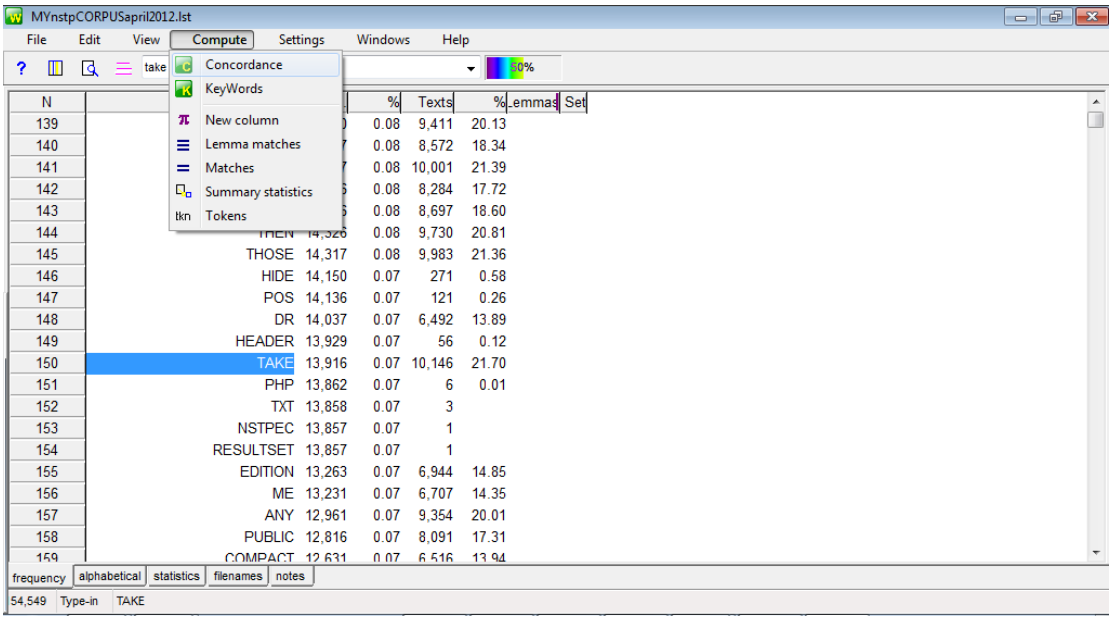
And then when the ***OPEN*** button is employed, the file or created corpus' name appeared. So, in this case the corpus named MyNSTPcorpus April2012 is selected.

Figure 3.11 Step 2 of the restricted collocations frequency search in NST corpus



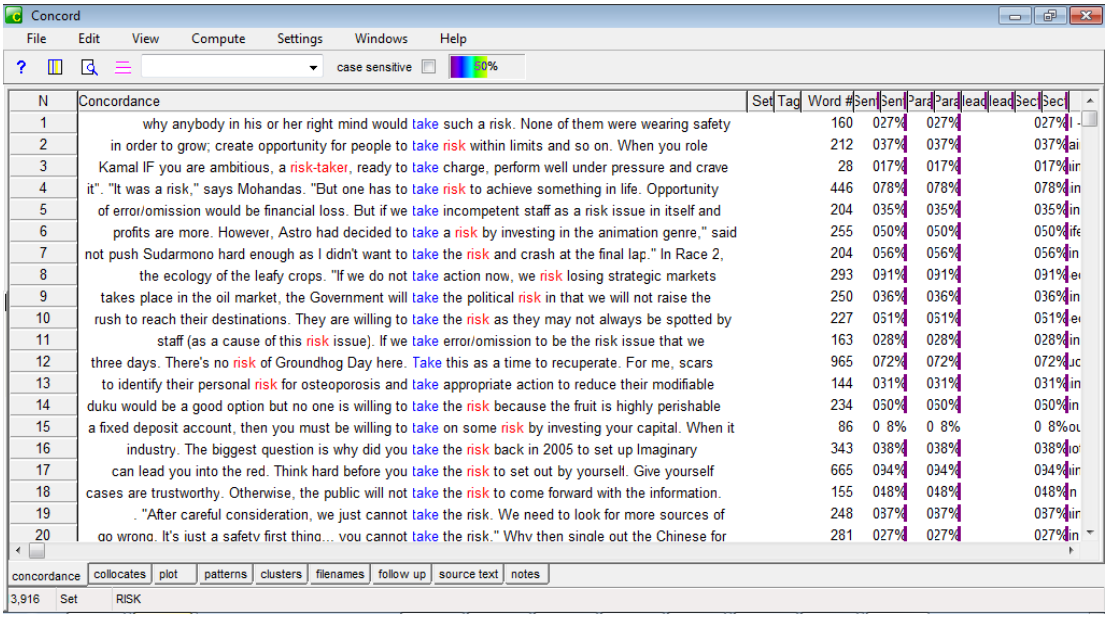
Once the file is opened, a frequency list of individual words appears, and a head verb is chosen for every single assignment needed for the concordancing program. A head verb is typed in the search column on the top as in the example in Figure 3.12.

Figure 3.12 Step 3 of the restricted collocations frequency search in NST corpus



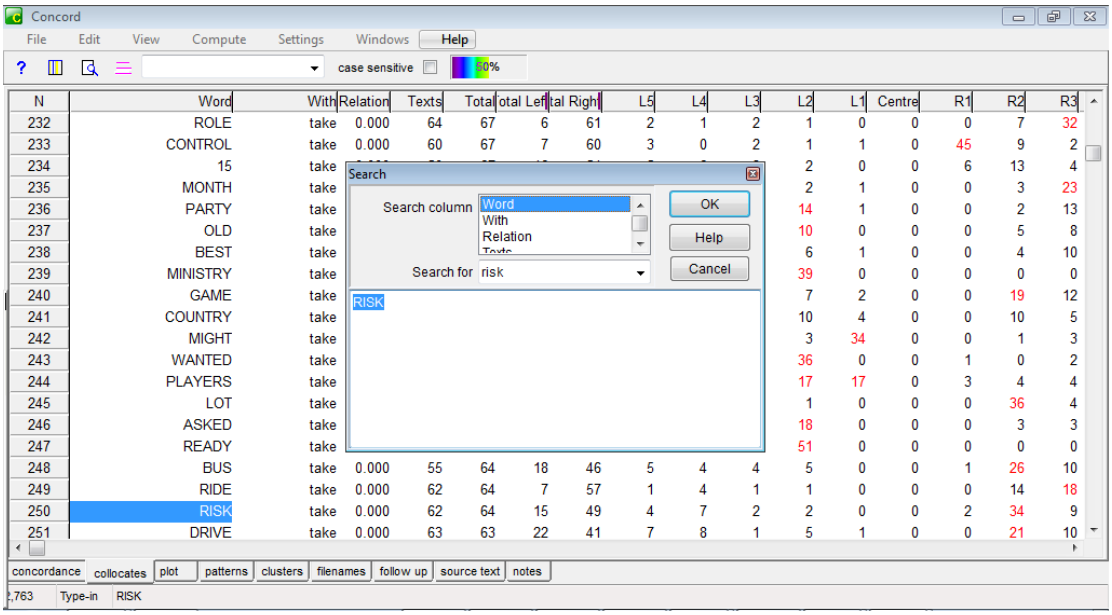
It takes few seconds for the result to appear. So the progress of the task is checked from the status bar progress which appeared in percentage. Figure 3.13 is an example of the search for the head verb *take*.

Figure 3.13 Step 4 of the restricted collocations frequency search in NST corpus



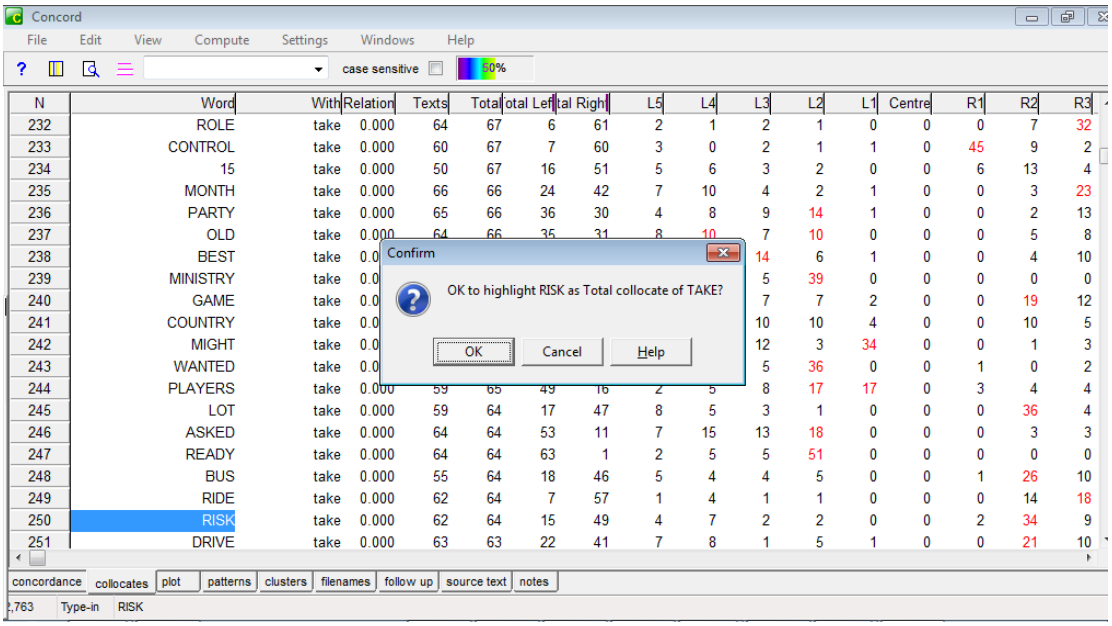
Then the collocation search is done to find the collocations of ‘take’ and ‘risk’. From the displayed window, the search result for the collocations will appear. The counting is done based on the existence of the string chosen. The number on the left pane made the counting easier to do.

Figure 3.14 Step 5 of the restricted collocations frequency search in NST corpus



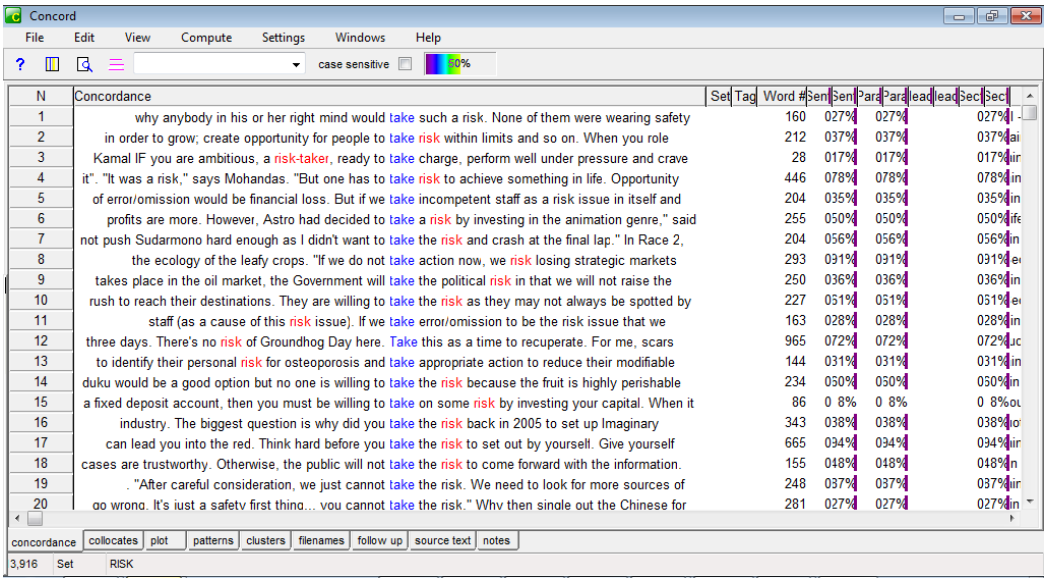
The figure below shows how the noun *risk* is found and highlighted so that a new window pops up the result of the collocations of *take* and *risk.*

Figure 3.15 Step 6 of the restricted collocations frequency search in NST corpus



The results that appear will be similar to Figure 3.15 above, where the collocations needed are highlighted. The following figure shows the search results for *take risk*.

Figure 3.16 Step 7 of the restricted collocations frequency search in NST corpus



The counting task is done with the aid of the highlighted words *take* and *risk*. Similar steps are applied to other verbs which share the same lemma. The same procedures are applied to all 20 sets of lexical verbs to be tested in the cloze test.

3.4.2 Summary of the corpus building

Below is a chart summarizing the process from corpus building to extracting the verb frequency list.

Figure 3.17 Summary of the process from corpus building to extracting the verb frequency list

Obtained the NSTP 7 years

of news archive

I

The news reports were copied into text files

(only 6 years; 2005-2010)

I

The news reports were filtered

(XBiz/sports news/jargons)

I

POS tagging using CLAWS tagger 7.0

I

Run data in Wordsmith Tools 5.0 to get the word list

I

A frequency list was obtained using Wordsmith Tools 5.0

I

The frequency list was checked against the frequency obtained from CELEX

I

A Frequency list was confirmed

I

20 head verbs based on bands were decided

I

20 strings or restricted collocations were shortlisted and checked for frequency in NST corpus and SAID list

I

The CLOZE test was developed based on the list

It is important to note that this corpus may not be an entirely accurate reflection of the full linguistic variation in the target population. However, it is 'specialized' and representative of a variety of an outer circle English, i.e. formal written Malaysian English. It is the only large source of evidence of its kind which currently exists. It may not be necessarily balanced but its data is sufficient for this project.

The following section will discuss the procedures of constructing the CLOZE testing items for Study 1 of the project and the other test items for Study 2.

3.5 Constructing the CLOZE Test

3.5.1 Introduction

The numbers of memorized fixed expressions or formulaic sequences are immense (Jackendoff, 1995). Interestingly nobody has been able to estimate the exact total number of these expressions stored in the mental lexicon of a native speaker. Assuming that their number could possibly even be larger than single words is worth suggesting, as Sinclair (1991) has gone so far as to argue that the dominant structuring feature of language is the *idiom principle*, rather than the rule based *open-choice principle*. Though these are diametrically opposed (Moon, 1998), both complement each other and are required in order to account for language use. This phenomenon provides a motivation as well as further evidence of how vocabulary is stored and retrieved.

It is important to use a good measurement tool when assessing the number of words learners have acquired. Tests that measure the learners' understanding of words when listening or reading are called receptive tests while tests which look at learners' speaking and writing are productive tests. There are different tests which can be employed when investigating L2 learners' collocational knowledge, use and development. And those elicitation tools have been used by researchers for years. Among the tests are multiple-choice and matching (and judgment) tasks (Granger, 1998; Gyllstad, 2007; Siyanova and Schmitt, 2008), recognition task (Gyllstad, 2007), and association task (Fitzpatrick, 2012). Another type of task is text chunking (Lewis, 1997; Boers, Eyckmans, Kappel, et

al., 2006; Eyckmans, Boers and Stengers, 2007; Stengers, Boers, Eyckmans and Housen, 2010). This task involves asking students to highlight or underline word strings in authentic texts, or the students may be asked to orally retell the content. Apart from essay writing (Li and Schmitt, 2010; Durrant and Schmitt, 2009; Howarth, 1998; Laufer and Waldman, 2011) and writing translation tasks from L1 to L2 (Webb and Kagimoto, 2011; Bahns and Eldaw, 1993), gap filling-tasks such as cloze tests and fill-in-the blank tests are among the regularly used tests to assess learners' collocational knowledge (Durrant, 2008; Durrant and Scmitt, 2010; Bahns and Eldaw, 1993).

As pointed out earlier, one of the tests which allows for the investigation of the retrieval of phrasal expressions is cloze testing. This is a common test which requires a reader to provide missing words in order to complete an expression on the assumption that context can trigger the retrieval of the expression from the mental lexicon. In addition to that, Stubbs & Tucker (1974) claim that a cloze procedure is not just a valuable test of ESL proficiency but also a tool which is quite powerful and easily used by a non-native teacher, and learners of English. Based on the *Superlemma model* (Sprenger et al., 2006; Kuiper et al., 2007), what is happening during speech production is that the activated superlemma node will in turn activate the lemma nodes of its constituent words. In this case, the reader is assumed to have knowledge of the expression if he/she is able to elicit or retrieve the missing words by filling up the gaps (Kuiper, Columbus, & Schmitt, 2009).

Cloze testing is quite a common tool used to assess language ability, especially in second language learning (Stubbs & Tucker, 1974; Alderson, 1979; Brown, 1989, 1993; Abraham & Chapelle, 1992; Kobayashi, 2002). Abraham and Chapelle (1992) and Brown (1989, 1993) detected that there was a close relationships between cloze item characteristics and item difficulty. Alderson (1979) states that some evidence shows that, the cloze procedure works better to test grammar and vocabulary than reading comprehension abilities. Apart from being easy to construct and run, this written measuring instrument entails high reliability and validity (Katona and Dörnyei, 1993).

There is a different view of research taken by Schmitt, Dörnyei, Adolphs & Durow (2004) where they 'have blended' the elements of cloze and C-test technique. The test item was designed to examine whether the participants could 'produce the formulaic sequence appropriate for the surrounding context' (p. 58). For that purpose, the context was left intact, and the content words in each formulaic sequence were deleted and a blank inserted. To help learners completing the words, the initial letter(s) of each word was given.

Some argue that the C-test is superior to cloze testing. However, the result of a study by Jafapur (1995) shows that C-testing may not be superior to cloze testing, contrary to what has been reported. This is due to its lack of validity. Other than that, the deletions do not affect a representative sample of the text which also suggests the invalidity of the procedure. However, there is a need to study in comparing these two testing tools. And for the present study, if C-testing is applied, where you give the first part of the missing

head-verb of a verb-noun collocation, it is considered you give the correct response away. This is due the fact that most of the head verbs used are short verb (get, does, make, give etc.).

To summarize, cloze testing is found to be an appropriate method for studies in Chapter 4 and 5 of the thesis. It is due to the fact that it is quite a common tool used to assess language ability in second language learning, and it is quite a standard procedure been used in the educational syllabus for English in Malaysia.

3.5.2 Objectives

There were a number of motivations for developing the cloze testing items:

a) A specific Malaysian English test set is required for Study 1 as the cloze items may not have the same frequency characteristics as Standard British or American English.
b) I am testing the acquisition of formal written Malaysian English, since that is the target for Malaysian student learners. Thus, an appropriate test item should be designed for this specific group of learners.

3.5.3 Procedure

The cloze items were selected based on their membership in four categories: light (or de-lexicalised) verbs (Grimshaw, 1990), non-light high frequency verbs, mid frequency verbs and low frequency verbs. These categories were determined based on the boundaries which have been drawn in Schmitt and Schmitt (2012) and Kuiper, Columbus & Schmitt (2009). However, the present study considers individual word forms rather

than word families (Nation, 2000; Adolphs & Schmitt, 2003) in the calculation of vocabulary coverage. Nagy et al. (1989) observe that learners seem to mentally handle the members of a word family as a group. Though there are pedagogic reasons to analyze vocabulary in terms of word families, it is impossible to program a computer to identify word families automatically. Among the limited software available for counting word families is Nation's RANGE program. This software functions based on referring to baseline lists of word family members which have already been compiled. Unfortunately the only possible way of obtaining reliable baseline lists is to do it manually, which is a very time-consuming method. Since the present study only focuses on verb+noun lexical collocations, i.e. verbs followed by their subcategorized complements, and the head verbs are the keywords, it seems that using individual words forms is sufficient. The words are grouped by *lemma*. Recall that a *lemma* is more restricted than a word family and includes only the base word and its inflections (Nation & Warring, 1997). So, for the purpose of this study, the individual lemma list was used. This was extracted from the NST corpus using the *Wordlist* program in *Wordsmith Tools 5.0*, since this computer-automated approach is more reliable in avoiding any possible errors in manual tabulation.

3.5.4 Categorization of verb frequency

For the purpose of both studies, the categorical list of verb frequency bands designed by Kuiper, Columbus, & Schmitt (2009) is adapted. The present study does not duplicate the frequency criterion, so only the bands and four categorical types are adapted. The categories were classified as shown in Table 3.3. This table also appeared in Section 2.6 (p.37).

Table 3.4 Kuiper, Columbus & Schmitt's verb frequency bands

Category	Frequency criterion
High frequency light verbs (HL)	Appearing in the top 1-3,000 words in the MFW lists (as words). Note that light verbs are also higher in frequency than the other high frequency verbs
High frequency lexical verbs (H)	Appearing in the top 1-3,000 words in the MFW lists (as words)
Medium frequency lexical verbs (M)	Appearing in the 3,000-5,000 word list in the MFW lists (as words)
Low frequency lexical verbs	Not appearing in any lists

The following table shows the frequency of the head verbs based on the Wordlist procedure from the lemmatized corpus using Wordsmith Tools, as well as the lemmatized NST corpus based on CELEX. The head verbs and the restricted collocations for Study 1 are shown in Table 3.4. They are in the appropriate bands and ordered by the frequency of the head verbs. The 20 items selected is probably too small a sample to be very representative. Yet, this could be slightly smaller than Cobb's (2007) target words which is 30 words (10 from each of the 1,000, 2,000 and 3,000). He was interested in whether vocabulary at various frequency levels occurred often enough to be learned merely from incidental exposure.

Table 3.5 Study 1 test items ordered by head verbs' bands

Head verb in each Phrasal lexical item(PLIs)	Frequency band
1. give a hoot	HLF

2.	making a fast buck	HLF
3.	get a grip of oneself	HLF
4.	does wonders	HLF
5.	taking a big risk	HLF
6.	stop bickering	HF
7.	pay respect	HF
8.	tell the difference	HF
9.	look the part	HF
10.	create a win-win situation	HF
11.	observe taboo and prohibited things	MF
12.	kill time	MF
13.	air view	MF
14.	cleared backlog	MF
15.	steal the show	MF
16.	shouldered the responsibility	LF
17.	rekindle family ties	LF
18.	gnash teeth	LF
19.	crack (a) joke	LF
20.	foot the bill	LF

Notes:
HLF- High Light frequency
HF- High frequency
MF- Medium frequency
LF- Low frequency

The above list of test items is obviously has several expressions that may be categorized as idioms by some phraseologists (e.g. kill time). There are possibilities that non-compositional phrases of such idioms may pose different challenges for learning (scoring in the test later) than the compositional ones (i.e. collocations). It is agreed that semantic transparency and also concreteness of meaning are important factors in incidental learning (Boers, Lindstromberg and Eyckmans, 2014).

The following table shows how the verbs were divided into 3 frequency bands; *high, medium* and *low* frequency bands with four categories based on a *combined evaluation* of verb frequency ranking in the NST corpus. The verbs were ordered by frequency. Refer

to Table 3.5 for Head verb frequency data and band allocation which were obtained using Wordsmith Tools 5.0, and Table 3.7 for Categorization of verbs.

Table 3.6 Head verb frequency data and band allocation obtained through Wordsmith Tools 5.0

Verb	Rank number (N)-frequency	Total occurrences	Frequency band
does	29	76 528	HLF
make	60	39 365	HLF
take	68	33 991	HLF
get	93	24 101	HLF
give	108	21 064	HLF
look	136	16 795	HF
tell	223	11 026	HF
pay	321	7 640	HF
stop	458	5 646	HF
create	492	5 283	HF
air	**3624 (*534)	643	MF
clear	769	3 538	MF
kill	1153	2 348	MF
steal	2182	1 205	MF
observe	3268	738	MF
crack	4084	550	LF
shoulder	**11 973 (*3017)	107	LF
rekindle	17422	56	LF
foot	**22 0598(*1339)	35	LF
gnash	69468	3	LF

Note
Rank number- *This figure includes the Noun category for the word 'foot', air and 'shoulder'
Total of occurrences- ** Verbs only' occurrences

Since the lemmatization in Wordsmith Tools 5.0 happened to occur with calculation and total up problems, the NST corpus was lemmatized by CELEX to avoid further problems. The table below shows a verb frequency listing extracted using CELEX as a lemmatizer. The frequency per million words is included as a way of normalizing the frequency scores. CELEX was used only to lemmatize the word forms, the tagged token counts derived from the NST corpus.

Table 3.7 Head verb frequency data and band allocation which obtained through CELEX lemmatization

Verb	Total occurrences	Frequency per million words	Frequency band
does	61 462	3,234.84	HLF
make	37 075	1,951.32	HLF
take	33 377	1,756.68	HLF
get	23 780	1,251.58	HLF
give	20 504	1,079.16	HLF
look	14 222	748.53	HF
tell	10 770	566.84	HF
pay	6 789	357.32	HF
create	5 218	274.63	HF
stop	4 261	224.26	HF
kill	1 997	105.11	MF
steal	913	48.05	MF
clear	907	47.74	MF
observe	728	38.32	MF
air	450	23.68	MF
crack	186	9.79	LF
shoulder	115	6.05	LF
rekindle	55	2.89	LF
foot	22	1.16	LF
gnash	3	0.158	LF

It is known that a small number of word types occur frequently and make up the majority of running words in discourse (Nation, 2006; Schmitt and Schmitt, 2012; Shin and Nation, 2008). What is obvious is that on how Nation treated the 3,000 most frequent word families is set to be the upper limit of high frequency vocabulary and that is representing the current best estimation of the basic learner lexicon of English. However, for this present study since there was a stage of difficulty of ranking the headwords (or base according to Nesselhauf, 2003) i.e. verbs, so the approach was to use the total of occurrences instead. To determine the frequency cut-off point for each band include several problems e.g. not too many verb-noun collocations found within particular bands.

It was then decided that best way in producing the cut-off points for the frequency bands was to combine the rank number (N) frequency of the verbs found using Wordsmith Tools 5.0, with total of occurrences from both CELEX and Wordsmith Tools. It might be questioned about the large gaps in the distribution of the high, medium and low. It was planned to replicate Nation's frequency bands where it was based on relative frequencies, e.g. the first 1000, and 2000 thousands etc. but the present study could not produce a 'neat' figures and distributions (similar to Nation's) across the range of the NST corpus.

Since the lemmatization was only for verbs so the bands or divisions are not in equal to other lists of frequency e.g. Nation (2006) or, Kilgarriff's (BNC) as there are only verbs were taken into account, and the corpus is too far too small and incomparable to others.

The selection criterion of verb categorization of NST corpus is listed below.

Table 3.8 Categorization of verbs in NST corpus

Category	Frequency criterion
High frequency light verbs (HLF)	Appearing in the top with the highest occurrences to 20,000 occurrences in the NST corpus list (as verbs only). Note that light verbs also tend to be higher in frequency than the other high frequency verbs. (Rank number (N) 1-1000verbs)
High frequency lexical verbs (HF)	Appearing in less than 20,000 to 5,000 words in the NST corpus list (as verbs only). (Rank number (N) 1-1000verbs)
Medium frequency lexical verbs (MF)	Appearing in less than 5,000 to 200 words in the NST corpus list (as verbs only). (Rank number (N)1000-3000verbs)
Low frequency lexical verbs (LF)	1-200 occurrences in the NST corpus. (Rank number (N) more than 3000verbs)

3.6 Checking the frequency effects of both verb frequency lists for Study 1 and Study 2

This section is going to discuss how the checking of the frequency was done for both word frequency lists designed for Study 1 and Study 2. The checking was done in order to validate whether both the frequency of head verbs and frequency of restricted collocations have effects on their occurrences in their respectively relevant corpora.

For this purpose there is a major thing to be observed: the relationship of head verbs frequency of both studies and the frequency of the PLIs selected for the cloze tests in their respectively relevant corpora: the NST (for study 1) and BNC (for study 2) corpora.

Before the test sets were administered to participants, the effect of the frequency of head verbs for both lists of PLIs were checked in their respectively relevant corpora: the NST and BNC corpora, to test Pawley's conjecture on head word text frequency and participation rates in PLIs. They were found to have a positive relationship, which will be presented later in this section. In other words, the higher the frequency of a verb, the more PLIs there are which have it as a head (Pawley, 1985).

The following results suggest that the higher the frequency a head verb has, the more likely a speaker will have come across a PLI with that verb as head. As a result, when guessing which verb might fill the cloze gap, a learner is likely to select one of the higher

frequency ones. This is especially true of idioms (K. Kuiper, personal communication, 2013).

A second analysis was done to investigate the relationship between the frequency of the head verb in the relevant corpus and the frequency of the whole PLI in the relevant corpus. A search in NST corpus using the Wordsmiths Tools 5.0 was done for the 20 restricted collocations selected for Study 1. Additionally, a search in the British National Corpus (BNC) was done for the 20 restricted collocations used in Study 2. The search using BNC was obtained through the web search interface (http://phrases inenglish.org/ searchBNC.html). A search for each of the 20 phrasal verbs was done in order to get the frequency of each set of PLIs. Phrases in English (PIE) is a database derived from the second or World Edition of the BNC (2000), but this database is not affiliated with the BNC Consortium. It provides an interface to search for words as well as phrases to a maximum of eight words long.

For this purpose, the '*Search the BNC for concordance'* was applied. This search provides an interface to query and return up to 1000 examples from the British National Corpus of the search terms highlighted in context. This query supports different kinds of matches; *the phrase*, *all the words* or *any of the words*. As the query *for all the words* provides wider matches, this option was best to opt for in this study. However, the frequency results were also checked against the *phrase* query. Figure 3.18 shows an example for the *phrase* type of query. The similarities and differences were matched and the results showed that the allocations were largely accurate. The query was better supported by the *wildcard operator* ***** which works in any position, including initial,

medial and final. Figure 3.19 illustrates an example of *wildcard* query. Together with this, the lemma search was also done in order to retrieve better results. Lemmatization in this case refers to the verb context such as *do*, *did*, *does* and *doing*. Figure 3.20 shows an example of the lemma results.

Apart from that, the search for the frequency of PLIs in NST corpus for Study 1 is not illustrated here, as the search is very similar to what is illustrated in Section 3.4.1. Basically, the searches for the 20 restricted collocations were quite similar for each string.

Figure 3.18 An example of a *phrase* type query using PIE

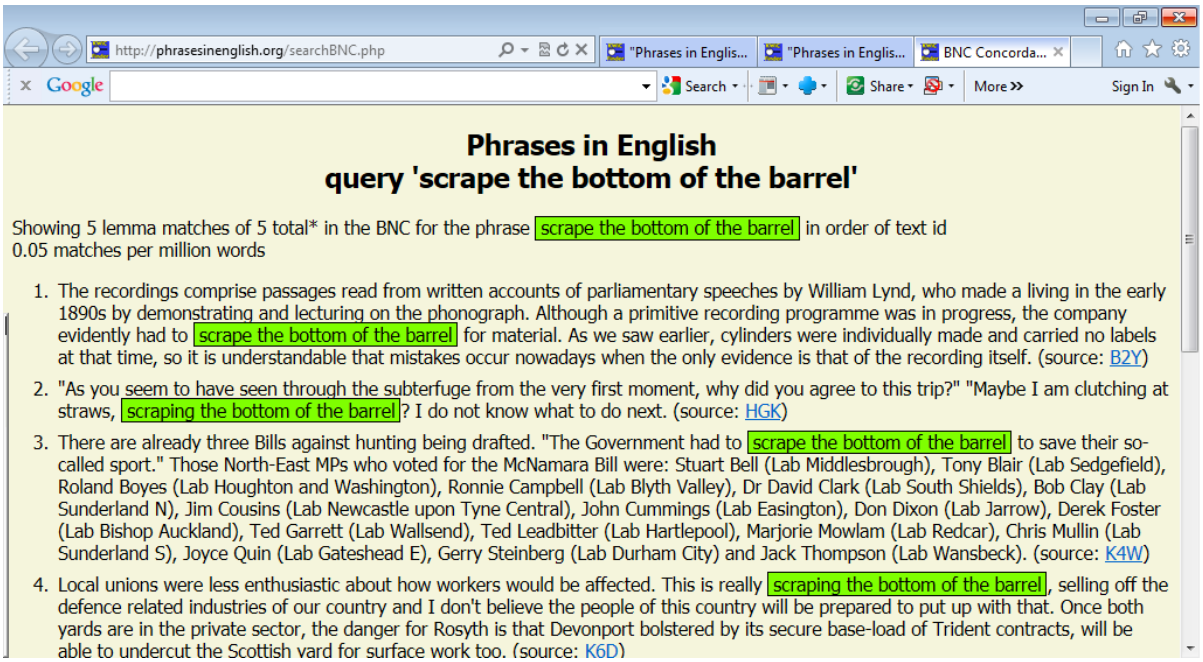


Figure 3.19An example of *wildcard* query using PIE

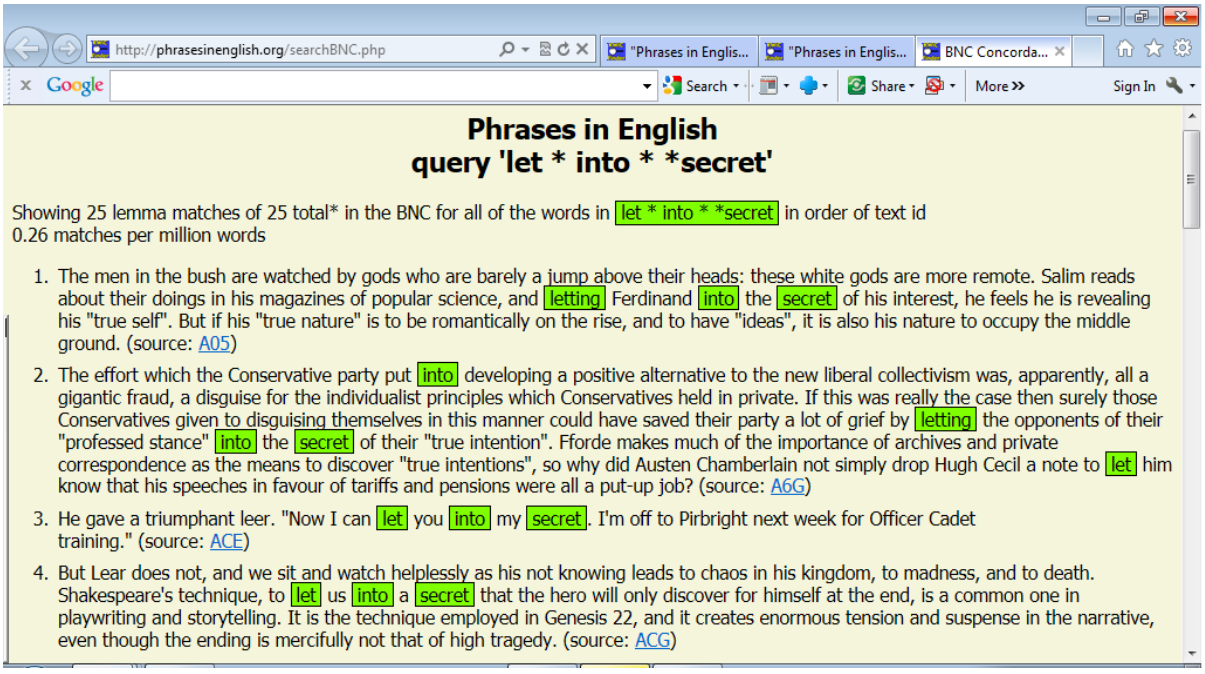
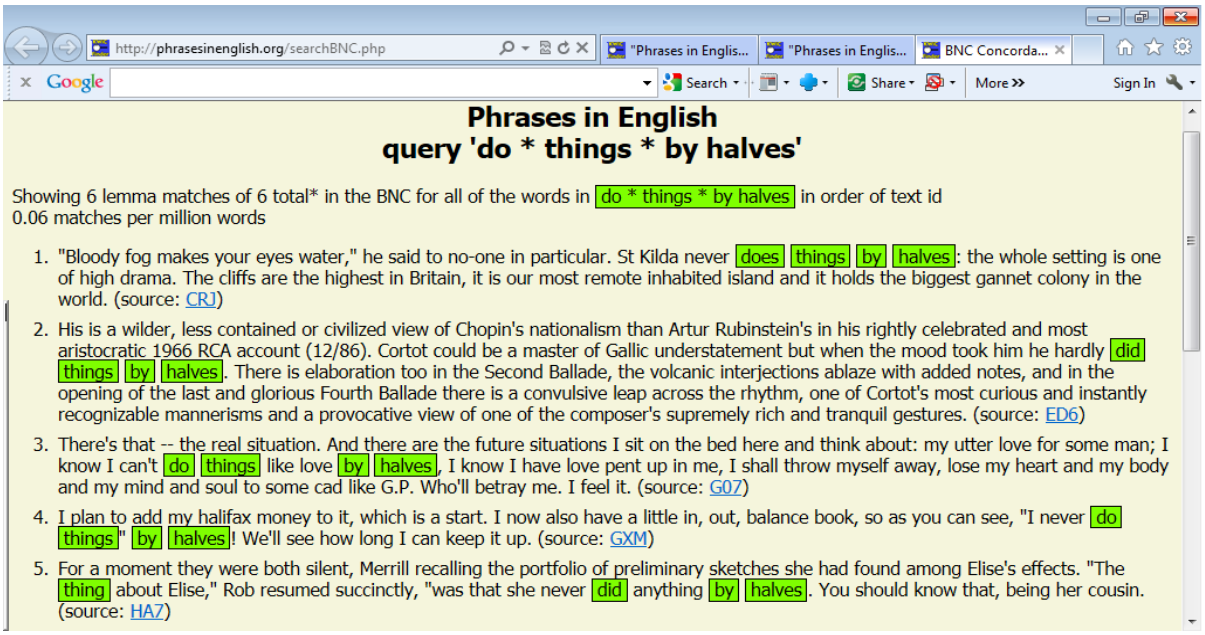


Figure 3.20 An example of *lemma* results from PIE



The frequency of the head verbs and the frequency of the restricted collocations based on their respectively relevant corpora are as shown in Table 3.8 and Table 3.9.

Table 3.9 The frequency of head verbs and the frequency of restricted collocations in NST corpus

Restricted collocations	Frequency band	Total occurrences (of head)	Frequency of (the exact) PLIs in NST corpus
does wonders	HLF	61 462	32
make a fast buck	HLF	37 075	20
taking a big risk	HLF	33 377	143
get a grip of oneself	HLF	23 780	6
give a hoot	HLF	20 504	3
look the part	HF	14 222	12
tell the difference	HF	10 770	27
pay respect	HF	6 789	99
create a win-win situation	HF	5 218	6
stop bickering	HF	4 261	7
kill time	MF	1 997	13
steal the show	MF	913	36
cleared backlog	MF	907	28
observe taboo and prohibited things	MF	728	2
air view	MF	450	31
crack(a) joke	LF	186	16
shouldered the responsibility	LF	115	39
rekindle family ties	LF	55	7
foot the bill	LF	22	27
gnash teeth	LF	3	2

Table 3.10 The frequency of head verbs and the frequency of restricted collocations in BNC

Restricted collocations	Frequency band	Total occurrences (of head)	Frequency of (the exact)PLIs in BNC
do things by halves	HLF	559 596	11
make tracks	HLF	217 268	31
take a fancy to	HLF	179 220	31
give NP the creep	HLF	131 417	30
keep a straight face	HLF	50 092	35
let NP into a secret	HF	29 768	23
join/enter the fray	HF	17 331	52

drive NP to drive	HF	16 477	8
act the goat	HF	15 620	6
avoid NP like the plague	HF	11 750	30
wipe NP off the map	MF	2 367	5
tighten NP's belt	MF	1 548	23
seal NP's fate	MF	1 512	16
spare no expense	MF	1 023	7
scrape the bottom of the barrel	MF	865	5
worship the ground NP walks on	LF	0	5
wring NP's neck	LF	0	24
pluck/summon up courage	LF	0	65
goad/spur NP into action	LF	0	28
toe the company line	LF	0	1

3.6.1 Evaluating the studies

Research Question:

The research question to be answered for this study is:

How do the frequency of the head verb in general and the numbers of PLIs with that verb as head in an idiom list (*SAID*) correlate?

The effect of the frequency of head verbs for both lists of PLIs were checked against the idiom list in *Syntactically Annotated Dictionary of Idioms (SAID),* a dictionary of idioms including phrasal verbs (Kuiper et al., 2003), to test Pawley's conjecture on head word text frequency and participation rates in PLIs. The correlation analysis was done only on the second data or Study 2 as the list of PLIs for standard English is available in SAID. The correlation result was found to have a positive correlation, which is presented in the following section.

Correlation between the number of PLIs with a certain head in SAID and the independent frequency of that head for data set 2. There is evidence of a relationship between these two properties (r=.523, n= 20, p=.018).

Table 3.11 The relationship between the frequency of the head verbs and the number of PLIs with that head in SAID

Correlation results of Study 2	
N	20
R	.523
P	.018

These findings show that there are high and positive relationships between the frequency of head verb in a corpus and the number of PLIs with that head in SAID. This suggests that Pawley's (1985) conjecture is supported.

The above study leads to another exploratory study in Chapter 6 of the thesis. That study is focused on individual acquisition of restricted collocations. So, it might suggest that when a respondent does not know a PLI they are more likely to insert a higher frequency verb in the cloze test for several reasons:

i. They are more likely to know the higher frequency verbs
ii. Since these verbs are associated with more PLIs they are more likely to choose the higher frequency verbs
iii. They are likely to choose light verbs because their meanings are so vague that they are likely to be a better guess if someone does not know an appropriate verb with a more definite meaning.

This will be further discussed in Chapter 6 of the thesis.

3.7 Testing the reliability of both cloze test items for Study 1 and Study 2

This section will briefly discuss how the testing of reliability of both cloze test question sets are done. There is a need for both cloze test tests to be tested for their reliability. A Cronbach Alpha was used to test the reliability of the two cloze set items utilized for the studies. Four participants from a group of postgraduate students from Malaysia had to answer both cloze testing sets of questions. Their score of both sets of testing tools were then calculated and checked for reliability scores. The reliability test checked for the four band (or categories) of both sets of cloze testing tools (for Study 1 and Study 2).

It confirms that both tests items are highly reliable, a= 0.915. Table 3.13 and Table 3.14 provide detailed results of the reliability test.

Table 3.12 Reliability results for both test items

Reliability Statistics	
Cronbach's Alpha	.915
N	8

Table 3.13 Individual reliability results for each item of verb frequency

	Scale mean if item deleted	Scale variance if item deleted	Corrected item-total correlation	Cronbach's Alpha if item deleted
L1	8.60	57.80	.985	.878
L2	9.20	79.70	.989	.913
HL1	7.60	58.30	.878	.891
H1	8.40	67.30	.797	.898
M1	8.00	55.00	.927	.887
HL2	7.20	69.20	.665	.909
H2	8.20	69.70	.924	.894

M2	8.60	87.30	-.013	.942

Note

L1	Low frequency (Cloze test 1-Study 1)
HL1	High (light) frequency (Cloze test 1-Study 1)
H1	High frequency (Cloze test 1-Study 1)
M1	Medium frequency (Cloze test 1-Study 1)
L2	Low frequency (Cloze test 2-Study 2)
HL2	High (light) frequency (Cloze test 2-Study 2)
H2	High frequency (Cloze test 2-Study 2)
M2	Medium frequency (Cloze test 2-Study 2)

This means that if someone scores (relatively) high on one category, they will score (relatively) high on all the others across tests (a=0.915). The two tests measure the same thing (but not necessarily on the same scale).

3.8 Chapter summary

Like other corpora projects employed to obtain word frequencies, the creation of the NST corpus falls into a number of complex phases yet came out with a very significant outcome to the overall project. Looking at the benefits of corpora I would agree with Kennedy's (1998) statement that 'linguists use corpora to answer questions and solve problems' (Kennedy, 1998: 2). Corpora provide us with authentic texts, and definitely linguists and researchers 'can sit at a computer terminal and call up all the examples of the usage of a word or phrase from many millions of words of text in a few seconds (McEnery & Wilson, 1996: 91). This illustrates the benefits of corpus data in language studies. I have a similar expectation that one day NST corpus will become useful to linguists, researchers and teachers in doing related tasks in language studies. However, there are some limitations of this corpus which need to be acknowledged. One of the weaknesses stems from the somewhat untidy nature of the corpus itself. However, there

are strengths as well to be highlighted and definitely this corpus will be further developed and eventually become balanced.

Apart from the weaknesses highlighted above, the state of this study using endonormative data, the NST corpus, rather than exonormative data for the overall study is actually an advantage and can be seen as an appropriate method of studying learners' vocabulary acquisition of other English varieties. The cloze test of the first study in this thesis is seen to be an essential methodology for testing in New Englishes. Thus, I am proposing that both corpus and cloze test used for this study have revealed a new kind of data for outer circle varieties where that variety of English is the target language.

CHAPTER 4
STUDY 1

The acquisition of Malaysian English restricted collocations

4.1 Introduction

This chapter reports the methodology and the analysis of the first study. This study focuses on the acquisition of Malaysian English restricted collocations. It examines the Malaysian English restricted collocations acquired by Malaysian learners both in Malaysia and New Zealand. The first study uses a newly developed cloze test where the verb frequency was supplied by the NST corpus. The details of the corpus development were discussed in Chapter 3 of this thesis.

In order to investigate the acquisition of Malaysian English restricted collocations, the test was administered to three different groups of Malaysian learners and a group of native speakers, as follows:

I. a group of high school students in Malaysia who were being taught English in a Malay medium system
II. a group of Malaysian undergraduate university students (from first, second and third years)
III. a group of Malaysian students in New Zealand (from first, second and third years)
IV. a group of native speakers of New Zealand English

The group of 20 native speakers of New Zealand English was recruited from among the undergraduates at the University of Canterbury. The rationale was to examine native speakers' responses towards the use of Malaysian English, as it was worth seeing how native speakers of New Zealand English responded to it to provide a baseline for native speakers. Their responses and reactions were essential, as Malaysian English was perceived as another variety of English. The results of their scores were seen to be a benchmark, and a way of checking the validity of this cloze test set.

This study was conducted through quantitative research methods, and analyses were done using SPSS software, which will be discussed below.

4.1.1 Research questions and hypotheses

The main research question and hypothesis for the study is:

How does the duration of exposure to the collocations of written Malaysian English affect the acquisition of Malaysian English collocations throughout the educational lifespan of learners of Malaysian English?

The hypothesis is that acquisition is positively correlated with exposure to RCs as measured by the length of time a learner has been exposed to the target language.

Other related research questions and hypotheses to be tested:

The frequency of a lexical item in corpora predicts acquisition because the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to the item. The following research questions were answered:

i) What role does the frequency of a clozed head verb of a verb phrase collocation play as a measure of the likelihood that a learner of Malaysian English has been exposed to a Malaysian English VP collocation?

The hypothesis to be investigated in answer to this question is that the frequency of the head verb in corpora of the target language is positively correlated with acquisition.

ii) What role does the frequency of restricted collocation as a whole, as a measure of the likelihood that a learner of Malaysian English has been exposed to a VP collocation, play on its acquisition in the case of Malaysian English? The hypothesis to be investigated is that the frequency of the RC in corpora of the target language is positively correlated with acquisition.

4.2 Methodology

4.2.1 Measures

According to Kuiper, Columbus & Schmitt (2009), each phrasal lexical item is a lexical unit with its own entry and it is stored in the mental lexicon of a language user. This theory, along with other relevant theories by Cutting and Bock (1997) and Sprenger, Levelt and Kempen (2006) looks at the nature of language acquisition. Specifically, they look at how lexical items are stored and retrieved, as well as what is acquired. It appears that cloze testing is considered as one of the tools that has the potential of allowing for the investigation of the retrieval of phrasal expressions to become possible (Kobayashi, 2002; Kuiper , Columbus & Schmitt, 2009). The cloze test designed for this study

consists of twenty gaps to be filled by the participants. Section 3.2 describes cloze testing in specific, providing a strong indication of how this test procedure is widely used in testing language ability (Kobayashi, 2002). The analysis looks at the participant's productive capacity for producing restricted collocations, and the extent to which they produce standard Malaysian English collocations. The acquisition of restricted collocations, measured by means of the cloze test, was analyzed according to the range of verbs which respondents use to fill the gaps. In most cloze testing (Kuiper, Columbus, & Schmitt, 2009), it is common to regard one of the possible cloze items as correct, i.e. idiomatic, and others as 'incorrect'. The former response is regarded as 'native-like' (Pawley & Syder, 1983) while the latter range of responses is regarded as 'non-native like'. In chapter 6, the aim is to see this as a continuum rather than a strictly binary matter.

4.2.2 Participants

Sampling procedures

The initial plan was to opt for a *clustered sampling* procedure, which involves choosing certain physical or geographical areas or states in Malaysia. This was supported by the latest data attained from the Ministry of Education in 2010, as there were about 242 registered private schools in Malaysia (see Appendix D). Among these schools only 64 schools were using English as medium of teaching and learning while the remaining 178 were using the Malay language or Bahasa Melayu as their medium of education.

Based on the given data, a number of schools were shortlisted to be chosen from urban and rural areas. The plan was to have at least two schools chosen from each area, making it a total of four or more schools, with two Malay medium schools and two English medium schools. This idea of having schools from various locations might have allowed for the observation of any significant effect of medium of instruction on the acquisition of restricted collocations. Unfortunately, the target was not possible to achieve as the schools' directors or principals were reluctant to either cooperate or join the project. This happened because almost all of the English medium schools are privately owned and those with authority, like the directors or principals, did not feel an obligation to participate in the research project.

At the initial stage, about 20 English medium schools were contacted, and only 2 schools from Gombak, Selangor agreed to join. Yet, it was later found that one of the schools used Malay as their main medium of teaching and learning. As for the Malay medium schools, which were government-based schools, there were two schools volunteered to join the research. Both of these schools were located in a state of the northern part of Malaysia known as Kedah.

In the end, only one English medium school participated in the study, with only limited numbers of students involved. With only seven participants, this group of students was too small and thus no longer significant for the study. As a result, the data collected from two Malay medium schools have become the only data from the school group.

Since there was very limited access to English medium schools, *convenience sampling* became the only available selection process. This sampling procedure was acceptable (Boudah, 2011) and this condition may suggest a limitation in generalizing the following results to a larger population.

Participants

The study was conducted at a number of institutions including two Malay-medium high schools in Malaysia. A total of 50 students from both schools were selected. Other than that, a group of 79 Malaysian university students were selected as participants. These university students were selected based on their year of education. For the next group, Malaysian students who are studying in New Zealand, a total of 53 students were chosen to become the respondents. All students who participated in this study answered the cloze tests, and prior to that they agreed to take part in the study by signing the consent form (see Appendix E). An information sheet regarding the study was also distributed to each participant (see Appendix E).

There were a total of 202 students including 20 native speakers selected for the cloze test and the details are as follows:

Table 4.1 Number of respondents involved in the study

Groups	Group details/no of students	Number of Groups	Total	Per cent %
High school students with non-English medium	School 1-30ss School 2-20ss	2	50	24.75

Malaysian undergraduate university students	Year 1-29ss Year 2-30ss Year 3-20ss	3	79	39.11
Malaysian students studying in New Zealand	Year 1-15ss Year 2-17ss Year 3-21ss	3	53	26.24
Native speakers of New Zealand English	Year 1,2&3-20ss	1	20	9.9
TOTAL			202	100

Note-
ss-students

All these students, except for the native speakers group, started their formal English education at the age of 7, which was the first year of primary school. The high school students were at the age of 17, which was the final year of secondary school education. As for the university students, they were different ages and came from variety of backgrounds, i.e. completed their A Level, completed their matriculation study, completed their diploma from polytechnics, or just completed their STPM-Sijil Tinggi Pelajaran Malaysia. The variety was already expected, but their primary and high school education are assumed to be standard. They are supposed to have attended 6 years of primary school education and at least 5 years in secondary school.

As noted above, there were 20 native speakers who were recruited for the study. These students were doing various courses at the University of Canterbury such as Psychology, Anthropology, Journalism, Geography, Commerce and Biological Science.
Tables 4.2-4.5 summarize the participants in each group.

Table 4.2 Questionnaire obtained from native speakers of New Zealand English

Native speaker of new Zealand English	No of students
Year 1,2 & 3 of undergraduate studies	20
TOTAL	20

Table 4.3 Questionnaire obtained from school students in Malaysia

School	No of students
Malay medium school	50
TOTAL	50

Table 4.4 Questionnaires obtained from university students in Malaysia

YEAR	No of students	Per cent %
Year 1	29	36.7
Year 2	30	37.97
Year 3	20	25.32
TOTAL	79	100

Notes
These students were doing various courses in a Malaysian university (UUM) i.e., Bachelor in Banking, Business and Entrepreneurships. Their gender is not counted for this study.

The following table lists the number of Malaysian students who are currently studying in New Zealand involved in the test. The students who participated were from various universities in New Zealand and doing various courses.

Table 4.5 Questionnaire obtained from Malaysian students who are studying in New Zealand

YEAR	No of students	Per cent %
Year 1	15	28.3
Year 2	17	32.08
Year 3	21	39.6
TOTAL	53	100

4.3 Procedure

The present study focuses only on verb+noun lexical collocations. As indicated above this is because they are regarded as key combinations in producing clauses and sentences, and they are the most often selected in the previous empirical research (Bahn & Eldaw, 1993; Bahns, 1993; Biskup, 1992). These studies also suggested that more focus is to be placed on verb+noun collocations, since it is the verb that causes the greatest difficulties for learners. Nesselhauf (2003) has the opinion that verb-noun combinations are the most frequently mistaken, so they should receive particular attention of learners.

The questionnaire and the cloze tests (see Appendix F) were distributed to each institution. Time taken for the cloze test was not controlled. The students answered the test with my supervision.

The following list consists of the 20 selected head verbs for the first study, divided into three frequency bands and four categories based on their frequency in the NST corpus. The first cloze test set used the restricted collocations as follows:

Table 4.6 List of 20 selected head verbs for the first study

Verb	Total occurrences	Frequency band
does	61 462	HLF
make	37 075	HLF
take	33 377	HLF
get	23 780	HLF
give	20 504	HLF
look	14 222	HF
tell	10 770	HF
pay	6 789	HF
create	5 218	HF
stop	4 261	HF
kill	1 997	MF
steal	913	MF
clear	907	MF
observe	728	MF
air	450	MF
crack	186	LF
shoulder	115	LF
rekindle	55	LF
foot	22	LF
gnash	3	LF

Note
HLF-High Light Frequency Verbs
HF-High Frequency Verbs
MF-Medium Frequency Verbs
LF-Low Frequency Verbs

The respondents were asked to fill lexical verb gaps from a text written by the researcher in a vernacular style. It was about a multicultural event celebrated in Malaysia. The aim was to maintain stylistic homogeneity throughout the task and provided sufficient

narrative interest to encourage respondents to maintain their interest until the end of the story (Kuiper, Columbus, & Schmitt, 2009). Given below are the 20 restricted collocations tested on the students. The students were only required to provide the missing verbs from the given text.

Table 4.7 List of 20 restricted collocations tested on the students

Restricted collocations
1. give a hoot
2. making a fast buck
3. get a grip of oneself
4. does wonders
5. taking a big risk
6. stop bickering
7. pay respect
8. tell the difference
9. look the part
10. create a win-win situation
11. observe taboo and prohibited things
12. kill time
13. air view
14. cleared backlog
15. steal the show
16. shouldered the responsibility
17. rekindle family ties
18. gnash teeth
19. crack (a) joke
20. foot the bill

4.4 Results

As mentioned in Chapter 1, the duration of exposure is interpreted as length of time a learner has been exposed to the target language. Since respondents were categorized into four different groups, they are expected to have different exposure, which should indirectly affect their performance on the test.

The subsequent sections discuss the two separate analyses using SPSS software. The *analysis of variance*, ANOVA was done in investigating learners' and native speakers' production of restricted collocations.

A correlation analysis was done on twenty restricted collocation to investigate the relationship between two variables, i.e. the frequency of *phrase frequency* as the independent variable, and a dependent variable, *acquisition*, which was represented by the total number of correct answers of only the learners. The Pearson correlation coefficients (r) take on values from -1 to +1. The motivation was to describe the strength and direction of the relationship between the variables. In this analysis the native speakers' group is omitted and only learners' total numbers of correct answers were counted as the focus was more on the production of restricted collocation of learners as native and non-native verb-phrase processing might not be the same. Plus, the aim was to look at the acquisition of Malaysia English and native speakers of New Zealand English might not be relevant to this analysis.

4.4.1 Participant groups and head verb frequency

Results

This section reports the results testing whether participant groups have an impact on the number of correct items. The hypothesis to be investigated in answering this question is that participant groups perform differently regarding test scores. Participant groups in this sense refer to the four different groups of participants involved in this study. It is also assumed that the frequency of a closed head verb of a verb phrase has effects on the total number of correct answers on the test.

Another hypothesis to be investigated in answer to this question is that the frequency of the head verb in a corpus of the target language affects the acquisition of RCs, which is represented by total number of correct answers.

For this purpose, a mixed between-within subjects analysis of variance (ANOVA) was done in order to examine the impact of participant group and frequency of head verb on acquisition, as measured by a newly developed cloze test. Remember that the four participant groups are Malay medium school students, tertiary students in Malaysia, Malaysian tertiary students in New Zealand and native speakers of New Zealand English.

The means and standard deviations are presented in Table 4.8 and Figure 4.1.

Table 4.8 Means and Standard Deviations for scores in the cloze test by 4 groups of participants and head-verb frequency category

Participant groups	Malaysian tertiary level N=79	Secondary school N=50	Malaysians in NZ uni N=53	Native speakers of NZ English N=20

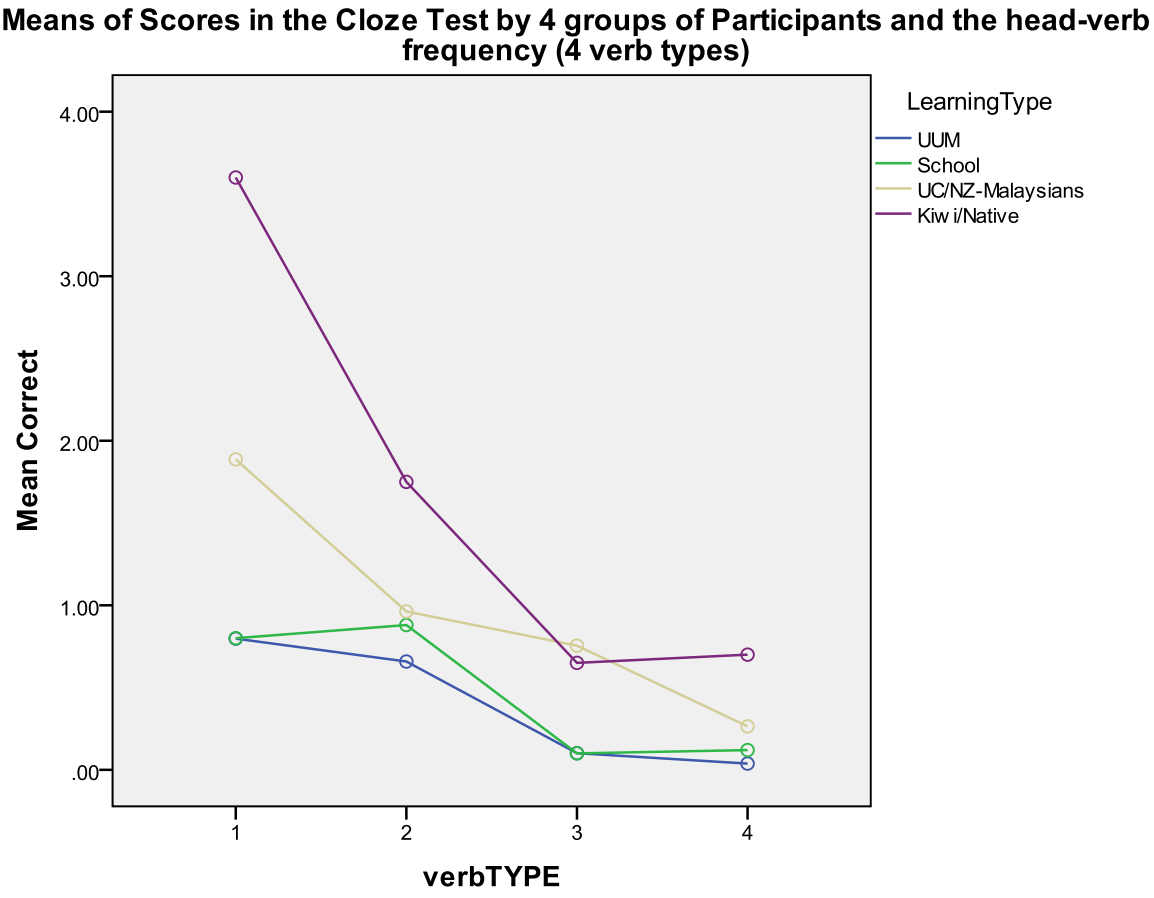
Verb type/Frequency	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation	Mean	Std Deviation
HLF	0.79	0.82	0.80	0.83	1.88	1.29	3.60	0.75
HF	0.65	0.57	0.88	0.59	0.96	0.61	1.75	0.85
MF	0.10	0.30	0.10	0.30	0.75	0.78	0.65	0.74
LF	0.03	0.19	0.12	0.38	0.26	0.52	0.70	0.73

Notes
The maximum score per frequency band (max= 5 per band).

HLF-High Light Frequency Verbs
HF-High Frequency Verbs
MF-Medium Frequency Verbs
LF-Low Frequency Verbs

Figure labelling:
Native speaker of NZ English - Kiwi/Native
Malaysians studying in NZ- UC/NZ Malaysian
Secondary school students-School
Malaysian tertiary level students- UUM

Figure4. 1 Means of Scores in the Cloze Test by 4 groups of Participants (Learning types) and the head-verb frequency (4 verb-types)



All pairwise tests were statistically significant ($p < .05$) with the exception of the difference between UUM (Malaysian tertiary group) and secondary school.

The interaction effect between four groups of participants and verb types was statistically significant, p= .000. There was a statistically significant main effect for verb type, p= .000 (less than alpha level 0.05), suggesting differences in the scores between the 4 different groups of learners.

The result between Malaysian tertiary group and secondary school group was quite unexpected as the tertiary level students were hypothesized to have more (longer) exposure than the secondary school students. They may share the same score for medium frequency verbs, but their scores for high frequency verbs (i.e. *stop, pay, tell, look and create*) and low frequency verbs (i.e. *shoulder, rekindle, gnash, crack and foot*) were slightly different. The effect of (length of) exposure is not observed across these two groups. It could be that exposure per se in not as powerful at all as explicit learning of collocations is much required and needed in this case. It seems that acquiring collocations through incidental learning is too slow (Gyllstad, 2007). What is more is that the gap of exposure is too small to show a gap and make a significant difference in terms of test scores.

The relative poor scores by the New Zealand native speakers were observed in supplying the missing head verbs. The participants responded well only in the first two verb frequencies i.e. high light frequency verbs, and slightly lower for high frequency verbs. The data showed that they performed poorly in the middle frequency and low frequency verbs. Since all the target collocations were considered standard and appear in BNC corpus, it could be that the vernacular style of the text which was written in the local Malaysian English style seem to be the main 'interference' of native speakers understanding the text. In other words, context may be one of the reasons causing difficulties for native speakers as they were not familiar with it, plus not enough clues to infer a correct verb for a specific collocation. Another reason could be that no credit was given to other appropriate collocates as the scoring approach was only considering the

exact-word scoring as correct answer. The exact-word scoring was an ideal approach which added objectivity to the marking process because it may not be affected by any kind of subjectivity or personal opinion (from the part of the evaluator). Alderson (1980) has stated that even native speakers (of a specific language) may not be perfectly correct, when the exact-word scoring approach is applied. However, the overall performance of native speakers was better as compared to non-native speakers. This was due to the RCs used were mainly used in standard English. They may face difficulties understanding the context (the Outer Circle English), but since they are familiar with English restricted collocations the retrieval is absolutely fast.

On a difficulty scale, expressions containing *high light frequency* verbs were considered the easiest ones, followed by expressions with *high frequency* verbs and *medium frequency* verbs, ending with the expressions including *low frequency* verbs, which were considered the hardest ones on the scale. As expected, the interaction effect between four groups of participants and verb types was statistically significant. There was also a statistically main effect for verb type. These findings suggest that both native speaker and three non-native speakers groups achieved a higher number of correct answers for the restricted collocations with high light frequency verbs than the restricted collocations using low frequency verbs, with the high and medium frequencies being in the middle. As can be seen in Figure 4, the effect of verb types (or verb frequencies) was observed as participants achieved the highest score on the most frequent category and followed by the other three categories or bands.

4.4.2 The role of frequency of restricted collocations on acquisition

Results

This section reports the results testing the hypothesis that the frequency of a restricted collocation correlates with the acquisition of Malaysian English collocations, which was measured by number of correct items on the cloze test. It is assumed that the frequency of the restricted collocations a whole correlate with the acquisition of RCs.

The frequency of a restricted collocation was interpreted as the frequency of each of the 20 restricted collocations tested in the cloze test. This frequency was based on the frequency in a corpus of a target language, i.e. NST corpus (see Table 3.8). The following table, Table 4.9 shows the summary of correlation results.

Table 4.9 Summary of correlation results for frequency 20 of restricted collocations and number of correct items on the cloze test by learners

N=20	
Pearson correlation of the frequency of restricted collocations and number of correct items on the cloze test	r = .035
p	.883

The results reveal that the relationship between the frequency 20 of restricted collocations and number of correct items on the cloze test by learners was too small. It shows a little evidence of a relationship between these two properties, (r = .035, N=20, p= .883).

The findings show that the there was too small relationship observed between the frequency 20 of restricted collocations (from NST corpus) and number of correct items on the cloze test by learners. In other words, the frequency of the restricted collocation as a whole does not play any effect on the acquisition of RCs of Malaysian English. Learners are not likely to answer the phrase verbs correctly even if the phrase has a higher frequency in a corpus. The findings suggest that phrase frequency is not a predictor of the acquisition of restricted collocations.

CHAPTER 5

STUDY 2

The acquisition of restricted collocations of New Zealand English

5.1 Introduction

This chapter will discuss Study 2, which shares quite similar procedures and test design with Study 1, presented in Chapter 4. This chapter will report the methodology and the analyses of the present research. It describes the research approach, sampling procedures, research instrumentation, data collection procedures, and the results of the analyses. Similar to the former chapter, the analyses utilized SPSS software and are discussed as follows.

The study focuses on the acquisition of standard English restricted collocations, in particular those of New Zealand English. This study is considered a replication study with Malaysian international students as respondents. It comes second because of the decision to look as students as they get further on in their studies and thus their exposure to English increased.

In order to conduct the second study, where learners acquire restricted collocations in a native setting, this study replicates the experimental design carried out in the study inquiring into the acquisition of phrasal vocabulary by Kuiper, Columbus, & Schmitt (2009). The test was administrated to 60 Malaysian students who were studying in New Zealand.

5.1.1 Research question and hypotheses

This chapter will cover the second research question of the study. The research question concerns the acquisition of restricted collocations of standard English by Malaysian learners abroad. The main objective of this study is to see the influence of exposure to standard English during the respondents' undergraduate studies. The learners' production of restricted collocations is examined throughout the immersion period, as it is suggested by Randall (1980), Schmitt (1998), and Sökmen (1993) that learners will acquire more 'native-like' phrasal vocabulary as their proficiency increases. Additionally, proficiency can only increase as a result of exposure to the target language over time, i.e. it is exposure grading. While this may be the case with many aspects of acquisition, it is an open question whether it is also the case with the acquisition of restricted collocations by international students. However, this study is not a longitudinal study as no data about the collocational competence of the participants before they came to study in New Zealand were collected.

The relevant research question and hypothesis to be answered are as follows.

Research question

How does the duration of exposure to New Zealand English affect the acquisition of standard English collocations in Malaysian international students?

Hypothesis

The hypothesis to be investigated in answer to this question is that the acquisition of RCs, or the number of correct answers, was different based on exposure, as measured by the length of time a learner has been exposed to the target language.

Other related research questions, and hypotheses to be tested:

It is assumed that the frequency of a lexical item in corpora is a factor in acquisition because the frequency of a lexical item is a proxy for the likelihood that a learner has been exposed to that item. This assumption results in the following research questions.

i) What role does the frequency of a clozed head verb of a verb phrase collocation, as a measure of the likelihood of having been exposed to a VP collocation, play in the acquisition of the corresponding RC in the case of standard English?

The hypothesis to be investigated in answer to this question is that the frequency of the head verb in corpora of the target language is positively correlated with the acquisition of the corresponding collocation in standard English or to be specific, New Zealand English.

ii) What role does the frequency of a restricted collocation, as a measure of likelihood of having been exposed to a VP collocation, play on its acquisition in standard English? The hypothesis to be investigated in answer to this question is that the frequency of the RC in corpora of the target language is positively correlated with its acquisition.

5.2 Methodology

5.2.1 Measures

The cloze test used for this study focused on the influence of exposure to English on the acquisition of English VP restricted collocations. This research study uses the experimental design of Kuiper, Columbus, & Schmitt (2009) which is outlined below. The rationale was that this group of international students was assumed to have adequate exposure to New Zealand English in their formal and informal activities. They were expected to be approaching the proficiency of native speakers and thus were expected to be able to perform similarly to the native speakers in the study by Kuiper, Columbus and Schmitt on test items.

As with the first study, the second experimental instrument also consisted of a vernacular story of a common social event (see Appendix G). According to Kuiper, Columbus and Schmitt (2009), at twenty points in the story there was a cloze gap where a verb was omitted. These verbs were classified according to their frequency in large corpora. The restricted collocations for this experiment came from the *Syntactically Annotated Dictionary of Idioms* (SAID), and the verbs were divided into three frequency bands and four categories based on a combined evaluation of verb frequency rankings in the Brown Corpus (accessed via edict.com.hk), Kilgarriff's BNC rankings (Kilgarriff, 1995), the Most Frequent Word lists (Nation, 2000) and the discussion in Nation and Waring (1997). These were checked against Kilgarriff's lemmatized BNC frequency list, and the results showed that the original allocations were largely accurate. The frequency bands were also checked against frequency data from the CELEX database and the BNC corpus data (www.natcorp.ox.ac.uk/). The choice of a cloze verb for respondents was open (not

multi choice). This test item was tested on participants aged from 16 to 60 plus. This was because there is previous evidence that the acquisition of phrasal vocabulary is age graded (Wray, 2002). Non-native speaker respondents were also tested using the same instrument because there was evidence that non-native speakers acquire phrasal vocabulary much more slowly than native speakers (Howarth, 1996; Wray, 2002).

5.2.2 Participants

Sampling procedures

For this research purpose, a stratified sampling procedure was applied for the study. A cohort of 20 students from each of Year 1, Year 2 and Year 3 of international study were chosen (N = 60). These students were Malaysian students studying in New Zealand, doing various courses in New Zealand universities, e.g. Bachelor in Mechanical Engineering, Geography, Sciences etc. Before the test started the students were given information sheet (see Appendix E) regarding the study. Those who agreed to take part in this study signed the consent form (see Appendix E). All students were required to answer the cloze test in a set up venue i.e. a hall. I supervised the test and there was no discussion allowed throughout the period. That was important as individuals' responses were crucial for this study.

Participants

Since the respondents were international students who were all from Malaysia and studying in New Zealand, they had acquired English as a second language, and the variety of English they were exposed to before coming to New Zealand was an outer

circle variety. This made for a very interesting population of international students, with exposure to both Malaysian and, more recently, New Zealand English. While there was a competency threshold for entry to New Zealand undergraduate study of IELTS 6, the variety of English with which students were most familiar was not standard British or North American English. (That is not to say that they have no exposure to these other varieties, through the media, for example.)

Most Malaysian students started their formal English education at the age of 7, the first year of primary school. The assumption made was that all of them started at the same age. However, these students were of different ages and came from a variety of backgrounds, i.e. completed their A Level, completed their matriculation study, completed their diploma from polytechnics and other tertiary institutions or just completed their STPM-Sijil Tinggi Pelajaran Malaysia, an examination equivalent to A Level. Some of these students were self-sponsored where they might come from well-to-do families, and some of them were on scholarship. This variety was expected, although their primary and high school education was assumed to be standard; most students attend 6 years of primary school education and at least 5 years of high school education in Malaysia

5.3 Procedure

This study, like Study 1, focuses on verb+noun lexical collocations, for the same reasons. The difference is only that the second study concerns the production of standard English while the previous one investigated the production of Malaysian English. So, in testing

the students' language production, the questionnaire and the cloze tests were distributed by the researcher to all 60 Malaysian students who were registered as undergraduate students in New Zealand universities. As with the first study, the time taken for the cloze test was not controlled.

The second cloze test set used the restricted collocations as found in the study by Kuiper, Columbus & Schmitt (2009):

Table 5.1 Frequency data and band allocation of head verbs

Verb	Rank number	Total occurrences	Frequency band
Do	18	559 596	HL
Make	46	217 268	HL
Take	54	179 220	HL
Give	76	131 417	HL
Keep	189	50 092	HL
Let	330	29 768	H
Join	594	17 331	H
Drive	618	16 477	H
Act	654	15 620	H
Avoid	866	11 750	H
Wipe	3122	2 367	M
Tighten	4178	1548	M
Seal	4249	1512	M
Spare	5457	1023	M
Scrape	6011	865	M
Worship	no rank	0 (as verb)	L
Wring	no rank	0	L
Pluck	no rank	0	L
Goad	no rank	0	L
Toe	no rank	0 (as verb)	L

Notes:
HL- High Light frequency verbs

H- High frequency verbs
M- Medium frequency verbs
L- Low frequency verbs

As for the cloze test, the respondents were asked to fill lexical verb gaps from a text in a vernacular style which was designed by Kuiper, Columbus, & Schmitt (2009). The complements of the lexical verbs being used were given in bold type. The aim was to provide a visual clue that the gap was related to the bolded sequence of words. The following table provides the list of the 20 collocational items used for the test:

The restricted collocations were ordered by band of head verbs as shown in Table 5.2

Table 5.2 List of restricted collocations used for Study 2

Phrasal lexical verbs (PLIs)	Frequency band
1. Do things by halves	HL
2. Make tracks	HL
3. Take a fancy to	HL
4. Give NP the creeps	HL
5. Keep a straight face	HL
6. Let NP into a secret	H
7. Join/enter the fray	H
8. Drive NP to drive	H
9. Act the goat	H
10. Avoid NP like the plague	H
11. Wipe NP off the map	M
12. Tighten NP's belt	M
13. Seal NP's fate	M
14. Spare no expense	M
15. Scrape the bottom of the barrel	M
16. Worship the ground NP walks on	L
17. Wring NP's neck	L
18. Pluck/summon up courage	L
19. Goad/spur NP into action	L
20. Toe the company line	L

Notes:
HL- High Light frequency verbs
H- High frequency verbs
M- Medium frequency verbs
L- Low frequency verbs

5.4 Results

The duration of exposure is interpreted as length of time a learner has been exposed to the target language. Other than age, other aspects like *year of study* and *months of learning* potentially affect the acquisition of restricted collocations. In this study, these three factors have been analyzed and discussed and termed as *exposure grading*.

i) Year of study ii) Age iii) Months of learning

The following sections discuss the three separate analyses using SPSS software. The analyses of ANOVA, Correlation and Multiple Regression analyses were done to investigate learners' production of restricted collocations throughout the immersion period.

A correlation analysis was done on twenty restricted collocation to investigate the relationship between two variables, i.e. the frequency of *phrase frequency* as the independent variable, and a dependent variable, *acquisition*, which was represented by the total number of correct answers of the 60 learners. The motivation was to describe the strength and direction of the relationship between the variables.

5.4.1 Year of study and head verb frequency

Results

This section reports the results testing the hypothesis that year of study has an impact on the acquisition of standard English collocations by Malaysian international students. The

hypothesis to be investigated in answer to this question is that the acquisition of RCs is affected by year of study.

It is also assumed that the frequency of a clozed head verb of a verb phrase collocation has effects on the acquisition of RCs. The hypothesis to be investigated in answer to this question is that the frequency of the head verb in corpora of the target has an effect on acquisition.

A mixed between-within subjects analysis of variance (ANOVA) was conducted to explore the impact of years of study of New Zealand English, represented by Year 1, Year 2 and Year 3, and the frequency of the clozed head verb of verb phrase collocations, represented by verb-type, on the acquisition of restricted collocations, as measured by a set cloze testing item.

The interaction effect between year of studies and verb types was not statistically significant, p= .205. There was a statistically significant main effect for verb type, p= .0001. However, the effect of year of studies was not statistically significant, p= .345, suggesting no difference in the score between the year of studies of those 60 students.

The means and standard deviations are presented in Table 1 and Figure 1. On a difficulty scale, expressions containing *high light frequency* verbs were considered the easiest ones, followed by expressions with *high frequency* verbs and *medium frequency* verbs, ending with the expressions including *low frequency* verbs, which were considered the hardest ones on the scale.

Table 5.3 Means and Standard Deviations for Scores in the Cloze Test by 3 groups of Participants (Year of Study) and the head-verb frequency (4 verb-types)

Year of Study(group)	YEAR 1 N=20		YEAR 2 N=20		YEAR 3 N=20	
Verb type/Frequency	*Mean*	*Std Deviation*	*Mean*	*Std Deviation*	*Mean*	*Std Deviation*
HLF	1.05	1.14	1.55	1.39	0.95	0.99
HF	0.70	0.86	1.10	0.78	1.15	1.13
MF	0.55	0.75	0.85	0.74	0.55	0.60
LF	0.30	0.65	0.30	0.47	0.25	0.63

Note

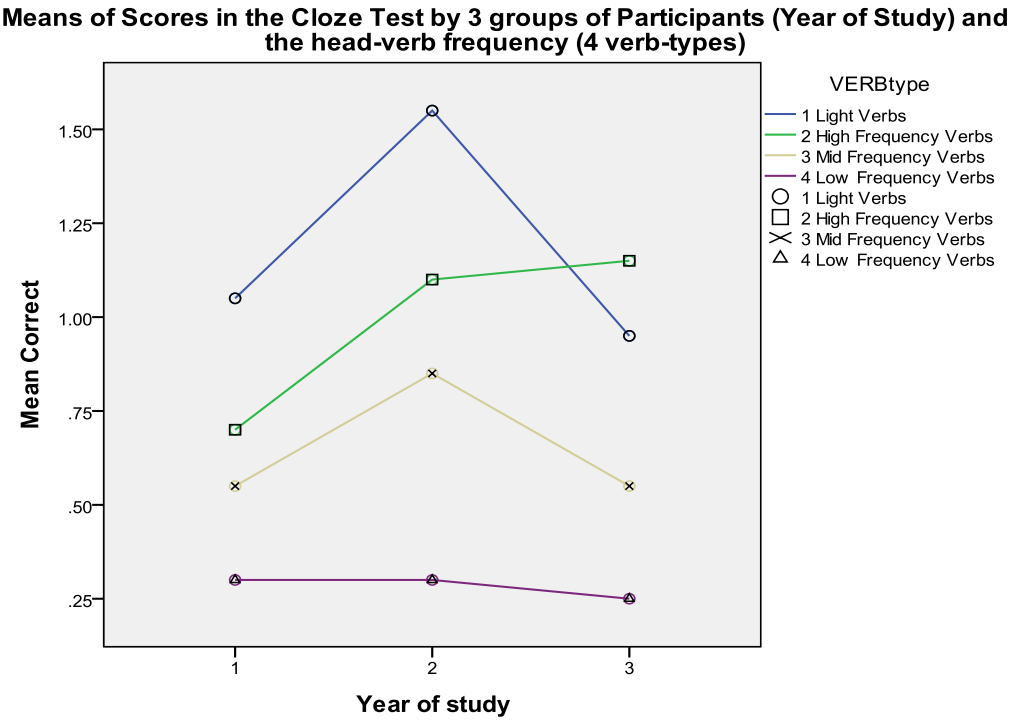
HLF-High Light Frequency Verbs

HF-High Frequency Verbs

MF-Medium Frequency Verbs

LF-Low Frequency Verbs

Figure 5.1 Mean of Scores in the Cloze Test by 3 groups of Participants (Year of Study) and the head-verb frequency (4 verb-types)



All pairwise tests were statistically significant ($p < .05$) with the exception of the difference between HFL and HF verb types.

5.4.2 Correlation results of year of study, age and months of learning

On the strength of these negative results relating to year of study, a further step which was not done in the analysis of Chapter 4 was taken by doing a further correlation analysis. The rationale was to observe the relationship among the variables, i.e. age, year of study and months of learning so as to confirm which one is a better predictor and can be used as a better *exposure grading* indicator.

Correlation analyses were done in order to describe the relationship among the variables. The Pearson correlation coefficients (r) take on values from -1 to +1. The variables of all four verb types: high light frequency, high frequency, medium frequency and low frequency evaluated along with other variables (total scores, age, year of study and months of learning) to describe the strength and direction of the relationship among the variables.

The results in Table 5.4 reveals that the relationship between months of learning and the verb types was strong, with a positive relationship between the variables (n=60). This means that more months of exposure led to a higher total score performance.

Table 5.4 The relationship between months of exposure and verb types

Head verbs	**r, relationship**
HLF	.283
HF	.300
MF	.380
LF	.300

Note
HLF-High Light Frequency Verbs
HF-High Frequency Verbs
MF-Medium Frequency Verbs
LF-Low Frequency Verbs

Apart from that there was a strong, positive correlation between months of learning and total score, r= .398, n=60. However, there was not a strong correlation between either year of study or age with other variables, especially total score and verb types, suggesting no strong relationship among these variable. The results are shown in Table 5.5. Thus,

months of learning may be a stronger indicator of the exposure grading, as it had a stronger relationship with the total score and the verb types.

Table 5.5 Correlation between either year of study and age with other variables especially total and verb types

		HFL	HF	M	L	total	age	year	month
	Pearson Correlation								
HFL	Sig. (2-tailed)								
	N	60							
	Pearson Correlation	.539**							
HF	Sig. (2-tailed)	.000							
	N	60	60						
	Pearson Correlation	.455**	.395**						
M	Sig. (2-tailed)	.000	.002						
	N	60	60	60					
	Pearson Correlation	.432**	.376**	.570**					
L	Sig. (2-tailed)	.001	.003	.000					
	N	60	60	60	60				
	Pearson Correlation	.851**	.780**	.731**	.694**				
total	Sig. (2-tailed)	.000	.000	.000	.000				
	N	60	60	60	60	60			
	Pearson Correlation	-.205	-.023	-.068	-.162	-.153			
age	Sig. (2-tailed)	.116	.864	.606	.216	.244			
	N	60	60	60	60	60	60		
	Pearson Correlation	-.034	.195	.000	-.035	.046	.571**		
year	Sig. (2-tailed)	.795	.134	1.000	.789	.727	.000		
	N	60	60	60	60	60	60	60	
	Pearson Correlation	.283*	.300*	.380**	.300*	.398**	.031	-.096	
month	Sig. (2-tailed)	.029	.020	.003	.020	.002	.816	.468	
	N	60	60	60	60	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Notes:

HLF-High Light Frequency Verbs

HF-High Frequency Verbs

MF-Medium Frequency Verbs

LF-Low Frequency Verbs

The means and standard deviations are presented in Table 5.6.

Table 5.6 Means and Standard Deviations of year of study and age with total score and verb types

	Mean	**Std. Deviation**	**N**
HLF	1.1833	1.20016	60
HF	.9833	.94764	60
M	.6500	.70890	60
L	.2833	.58488	60
Total (score)	3.100	2.68518	60
Age	21.13	1.478	60
Year	2.0	.823	60
Months	188.95	35.990	60

Notes:
HLF-High Light Frequency Verbs
HF-High Frequency Verbs
MF-Medium Frequency Verbs
LF-Low Frequency Verbs

5.4.3 Multiple regression analysis of year of study, age and months of learning

A multiple linear regression analysis was done to examine the relationship between the dependent variable, total score, and a set of independent variables or predictors. All the predictors were entered in a single step using a forced entry method. This method was opted for because in this method each variable was evaluated for its predictive power,

and this study aimed to evaluate how much unique variance in the dependent variable (total score) can be explained by each of the independent variables. In other words, no variables were controlled and each variable was expected to make a unique contribution. In this sense, the analysis tested the hypothesis of the influence of exposure to the acquisition of verb-phrase collocations. The exposure was referred to the exposure to the target language over time, i.e. it is exposure grading. The force entry method was also opted in testing the above theory of age grading in the acquisition of restricted collocations as in the studies done by Kuiper, Columbus & Schmitt (2009) and Escaip (2008).

This regression analysis was done to observe the relationship among those independent variables; age, months of learning and age, and dependent variable; total scores as the targets were to explore the following aims:

1. how well this set of independent variables are able to predict a particular outcome, which is the total scores,
2. to find out which variable in this independent variable set is the best predictor of the total scores

In this analysis verb types were not used as part of the dependent variables as the aim was not on the frequency effects in the acquisition of restricted collocations. The correlation results in Section 5.4.2 showed that months of learning has shown a strong and positive relationship among other variables i.e., age and year of study where more months of

exposure will result to higher score performance. Months of learning was seen to be a stronger indicator of the exposure grading.

Analysis of variance (ANOVA) and regression analyses may seem to be redundant in terms of finding out the effects of exposure grading on acquisition; however, the regression analysis seems to able to provide a clear answer as to which variables predict the outcome score. At the initial stage, ANOVA was done with the assumption that the acquisition of restricted collocations was affected by year of study (age graded), parallel with Kuiper, Columbus & Schmitt (2009) and Escaip (2008). Since, the year of study did not show an effect then multiple regression analysis was opted to figure out a better predictor. The main aim was to figure out a strong indicator of the exposure grading.

Preliminary analyses were conducted to ensure that there were no violations of the assumptions of normality, linearity, multicollinearity and homoscedasticity.

The model as a whole was 23.4%, Sig. Of F change=0.002. The model showed that only one measure, months of learning, was statistically significant, recording the highest beta value (beta= .433, p= .001). This is greater than those of age (beta= .321, p= .029) and year of study (beta= .270, p= .065).

This analysis has shown that months of learning made the strongest unique contribution to explaining the dependent variable, the respondent's total score, as compared to year of

study and age. This suggests that months of learning contributes the most to the prediction of acquisition of restricted collocations.

Table 5.7 The mean score and standard deviations of dependent variable, total score and a set of independent variables or predictors: age, year of study and months of learning

	Mean	Std. Deviation	N
Total score	3.1000	2.68518	60
Months	188.95	35.990	60
Year of study	2.00	.823	60
Age	21.13	1.478	60

The following is the model summary of the dependent variable, total score and a set of independent variables or predictors: age, year of study and months of learning.

Table 5.8 Predictors of total scores

	Model 1
Variables	
R	.484
R square	.234
Adjusted R square	.193
Std. error of the estimate	2.41240
R square change	.234
F change	5.699

df1	3
df2	56
Sig.	.002

Notes:
Predictors: (Constant: age, months, year of study

5.4.4 The role of the frequency of a restricted collocation on its acquisition

Results

This section reports the results testing the hypothesis that the frequency of a restricted collocation correlates with the acquisition of standard English collocations, which was measured by number of correct items on the cloze test. It is assumed that the frequency of the restricted collocations a whole correlate with the acquisition of RCs.

The frequency of a restricted collocation was interpreted as the frequency of each of the 20 restricted collocations tested in the cloze test. This frequency was based on the frequency in a corpus of a target language, i.e. BNC (see Table 3.9). The following table, Table 5.9 shows the summary of correlation results.

Table 5.9 Summary of correlation results for frequency 20 of restricted collocations and number of correct items on the cloze test by learners

N=20	
Pearson correlation of the frequency of restricted collocations and number of correct items on the cloze test	r = .147

p	.538

The results reveal that the relationship between the frequency 20 of restricted collocations and number of correct items on the cloze test by learners was too small. It shows a little evidence of a relationship between these two properties, (r = .147, N=20, p= .538).

The findings show that the there was too small relationship observed between the frequency 20 of restricted collocations (from BNC corpus) and number of correct items on the cloze test by learners. In other words, the frequency of the restricted collocation as a whole does not play any effect on the acquisition of RCs of standard English. Learners are not likely to answer the phrase verbs correctly even if the phrase has a higher frequency in a corpus. The findings suggest that phrase frequency is not a predictor of the acquisition of restricted collocations.

CHAPTER 6

The Analysis of the selection of non-idiomatic heads

6.1 Introduction

In this chapter I shall look ahead in a way that is often suggested with the questions: Where does this research lead? What might one do next? The research up to now has focused on groups. The results are cohort results. They tell us about means in an attempt to determine the effect of a number of variables on second language learners' acquisition of phrasal lexical items, specifically restricted collocations. In this chapter I outline an approach to acquisition which is focussed on individual acquisition. It does this by recommitting some of the data analysed earlier and coding it more delicately. Previously, the coding was essentially binary: a respondent either knew the restricted collocation or they did not. However, in this chapter the coding is different: the respondent knows the idiomatic restricted collocation or, if they do not, they enter an alternative which makes good sense in the context to a degree, ranked on a Likert scale from 1 to 6. Less plausible responses might be, for example, either entering the wrong syntactic category, i.e. a non-verb, or entering a verb that does not suit the context. The latter suggests that the respondent either does not understand the context sufficiently to enter an appropriate verb or their vocabulary knowledge is so slight that they do not know a verb which would fit the context.

The motivation for proceeding with an analysis in this way is that one might suppose that the acquisition of phrasal vocabulary occurs after earlier single word vocabulary is acquired, and that it is based on more extensive exposure to the target language. So we might suppose that a respondent who is faced with a lexical retrieval task, which is

essentially the nature of a cloze test, but who does not understand the context either semantically or syntactically, will opt for a wild guess. A more advanced learner who does understand the context and has acquired sufficient vocabulary will be able to fill the cloze gap with a semantically and syntactically appropriate filler. Finally, as learners become more proficient in the target language, they will more often know and enter an idiomatic filler. To assist with the recoding of the range of cloze test responses, 10 native speakers were assigned to rate the full range of individual responses. These native speakers of English evaluated the responses and coded them based on their acceptability in the context.

It is not the aim of this chapter to repeat all of the previous research with this more complex coding but to show, for one cohort (one year of study two), how such a study might be undertaken and what the results might look like. This will provide an account of all the responses produced for every question by the whole cohort of respondents. It will also then show the range of individual differences in each member of the cohort. The hypotheses to be investigated here are that:

1. Individuals who have higher scores of idiomatic responses will also have higher rates of plausibility for their non-idiomatic responses.

2. Of the non-idiomatic responses more will be at high end of the frequency spectrum.

If these investigations appear worthwhile, then they suggest clear further directions for research in that all the remaining data from both studies in Chapters 4 and 5 can be

similarly analysed to see if mean changes in performance, with increased exposure to the target language along the two stage path suggested here, actually takes place. Further research might also be conducted to follow a cohort of students through their years of exposure to the target language to see if individuals go through these two stages of acquisition and what the variability among individuals might be.

Individual cloze test items can also be evaluated using the above outlined strategy for how they discriminate amongst respondents by looking at how the analysis allows for discrimination among the cloze items themselves. Some items may be 'harder' both in terms of how many idiomatic responses they elicit as well as how the non-idiomatic responses rank in plausibility. Others may be 'easier' in terms of eliciting idiomatic responses, i.e. the idiom is well known, but if it is not, then the responses may be at the less plausible end of the plausibility scale. This suggests that cloze testing can be made sensitive to individuals and that the careful choice of cloze items can elicit better evidence of vocabulary learning than just coding responses as idiomatic or non-idiomatic.

Such an approach is congruent with the work of Dornyei and Skehan (2003), as well as Sawyer and Ranta (2001), who highlight the finding that individual difference factors have been shown to have significant impact on language learning in general. In addition, Schmitt, Dornyei, Adolphs & Durow (2004) argue that it is quite logical that they might also influence the acquisition of formulaic language.

Apart from supporting the second hypothesis, three learners' profiles are presented in detail here, as a means of demonstrating the opportunities for future research that combines single case studies with broader cohort profiles.

6.2 Background

6.2.1 The relationship between the exposure to a lexical item and the learning of that item

The major hypothesis of this current research has been that exposure to the target language seems to enhance the progress of vocabulary acquisition in that language, parallel to Fitzpatrick's notion (2012). As noted earlier, Fitzpatrick's study was conducted in order to understand how exactly the experience of studying abroad impacts the sophistication of word knowledge and lexical organisation. It revealed a gradual increase only in some aspects of vocabulary knowledge, while there were some aspects which were inconsistent. The results indicate that there is a relationship between word frequency and the stages of producing orthographic form. When the 'native speaker-like' responses were evaluated, there was a slight upward trend over the period of the study. Another interesting finding was that participant's knowledge of collocations seemed to elevate as the year progressed. Knowledge of the correct form became consistent after the third stage of the study. The finding also suggested the value of repeated exposure to lexical items. For this particular study intensive exposure and practice seemed to 'hurry' the progress of vocabulary acquisition. This finding suggests that what is needed is a much more nuanced notion of exposure.

That is why we need to look at cloze responses in terms of the target language resource which learners have access to. Learners will always fill a cloze item with what they think of as a plausible filler. But it may not be the most native-like filler. So how much exposure do learners have to a particular item of phrasal vocabulary? If acquisition is related to exposure, it must be exposure to particular vocabulary items, and many items

of phrasal vocabulary have a very low frequency of occurrence even for native speakers (Wray & Perkins, 2000; Wray, 2002; Peters, 1983).

The process being assumed here is as follows. A reader reads the cloze passage story. The story creates a context for what comes after. It does not predict what comes after but creates some 'stable' expectations. The slot is syntactically a verb, given that a subject comes before and a complement after. So, in this case the reader needs to 'supply' what is left unsaid (Halliday & Hassan, 1976). This is one of the reasons why V' collocations are selected for testing.

For this research purpose, the task was designed for learners to retrieve from the mental lexicon either an idiomatic filler or, if that was not known, a semantically plausible verb which fits in the slot. This sort of gap filling task is not primarily a perception task but rather concerns production. If this production task induces learners to fill the gap with a unique item in native speaker norms such as those in restricted collocations, rather than a semantically plausible word, then this is evidence of this lexical item being accessed as a whole from the speaker's mental lexicon. Thus, this state is closely related to the theory of spreading activation. According to this speech production theory, word retrieval requires selecting a lemma, a lexical representation that is semantically and syntactically specified, from all other lemmas stored in one's mental lexicon, followed by phonological encoding of that lemma (Dell, 1986; Levelt, 1989). So, in this case, a word is selected if its activation exceeds some threshold, otherwise the most activated word is opted for after a fixed period of time (Dell, 1986). However, if that filler has not had

sufficient activation in the past as part of a phrasal lexical item, or in other words there has been insufficient exposure to it in the past, then the learners need to find something that fills the gap plausibly. If they were very uncertain in their language knowledge, then they might just put anything that comes into their head into the slot. What comes into their head must then be the result of word associations (Fitzpatrick & Izura, 2011).

The following subsection clarifies the basis of why 10 native speakers were recruited to help with the coding. Their discretion was essentially vital to assist with the recoding of the range of cloze test responses. These native speakers' evaluation of acceptability was seen as unique, as they treat phrasal lexical items quite differently from non-native speakers.

6.2.2 Native and non-native collocational knowledge

Both Howarth (1996) and Granger (1998) studied the use of collocational sequences of native and non-native writers of English. Both studies found significant deviations from standard native norms made by the learners and, on top of that, suggested that learners did not approach the phenomenon from the same directions as native speakers. The deviations from such standard norms were traced based on the errors made as well as the fact that such sequences were used less frequently by learners. While learners may avoid using these collocational sequences due to lack of knowledge, native speakers may see this as *shortcuts* (Peters, 1983: 82; Hickey, 1993: 29; Wray, 2002: 106), as native speakers have been exposed to many of these sequences in their input since they were young (Wray, 2002). This has shown that though collocations are highly significant in communications for both learners and native speakers, apparently they are closely related

to extensive exposure to a target language which in this case is English. While learners may struggle, native speakers can fluently produce multi-clause utterances because many constituents of them are memorized as prefabricated phrases (Pawley & Syder, 1983).

This fact led to the decision of recruiting 10 native speakers for the coding purpose. They rated the learners' responses based on the acceptability and plausibility.

6.3 Methodology

6.3.1 Participants

The study participants were 20 final year undergraduates from the University of Canterbury. The participants were among the 60 respondents in Study 2, discussed in Chapter 5 of this thesis. As mentioned earlier, they shared the same background; these were final year students, ranging from 17 to 23 years old, and they were doing various courses, e.g. Engineering and Geography. Before the data for this study were collected, they passed IELTS (International English Language Testing System) with a minimum of Band 6. This test is an international standardised test of English language proficiency. It is jointly managed by the University of Cambridge ESOL Examinations and the British Council. The participants' results were used as a prerequisite to studying in New Zealand. Participants' native languages were Malay, Tamil, Mandarin, Cantonese or other Chinese languages. They had been learning English in school from age 7, and had at least 11 to 17 years of overall exposure to English. These students were sharing accommodation with other Malaysian friends since there is a reasonably large Malaysian student community in the university campus.

As mentioned earlier 10 native speakers with linguistics background were recruited for the purpose of coding. Their evaluations were used to rate the acceptability of the responses provided by the learners. The highest score of 6 was given if the response was entirely acceptable while the other extreme of 1 was graded if the response was considered as completely unacceptable.

Among the 20 participants, one learner's profile is presented as a means of providing detailed documentation of an individual's personal lexical knowledge based on the cloze test results. This learner's non-idiomatic responses were listed and coded and also checked for verb frequency rank with Kilgarriff's lemmatized BNC frequency list. Apart from that, two other learners' results are presented and compared with the individual's profile. The motivation for proceeding with this analysis was to test the second hypothesis with the presumption of the use of more verbs at the high frequency end of the spectrum if less exposure was received. The female participant is a native speaker of Malay and was 23 years old during the data collection period. This participant, who used the nickname *Anna*, was a final year Mechanical Engineering student at the University of Canterbury. She passed the IELTS with Band 6 and has been learning English for 17 years. She attended government-based school in Selangor, Malaysia, and had a year preparation course in Malaysia before coming to New Zealand. She was also sharing a house with 2 other Malaysian friends though she realized that might not benefit her English language development. However, she was a highly motivated learner. Apart from studying hard for her Mechanical Engineering degree course, she loved language learning and she was quite fond of English literature, English movies and reading novels. This was an aspect that made her different from the majority of the other participants. As suggested

by Dörnyei, Durow, & Zahran (2004), her attitude towards English might have some kind of connections with her language performance and might account for her collocation production too. The other two learners' results were analyzed for comparison. These two respondents could be examples of those with less exposure and this would be useful to answer the second hypothesis. Their detailed profiles may not be presented comprehensively as their results were the major reference for the analysis.

6.3.2 Procedures

All the responses of the test cohort of respondents on each of the twenty cloze items are listed as in tables 6.2 – 6.21. They were then reclassified. The idiomatic responses had already been selected in study 2 (Chapter 5); therefore the task was to take the non-idiomatic responses, i.e. the full list of responses used by the respondents which were not the idiomatic one(s), and sub-classify them according to how plausible native speakers perceived them to be. As mentioned earlier, there were ten native speakers who were asked to perform this classification. They were asked, 'Does the insertion of each of the following words into the story at this point make good sense or not?' Their task was to place a score of 1-6 in the provided column by indicating how acceptable they found the word in the given context. Their scores:

1- completely unacceptable- I can't imagine this word being used in this context
2-
3-
4-
5-
6- entirely acceptable – I would use this word in this context

They were informed that the most obvious answer may be missing from the list and they were required to rate each word as it fits in the gap. The mean value and standard deviation were then calculated across the responses in the two new categories. Provided below is the list of the verbs used for the analysis. Note that the verbs are specific to each cloze gap.

Recall that the phrasal lexical items used in the cloze test are shown in Table 6.1 (This table is also to be found as Tables 3.9 and 5.2)

Table 6.1 List of phrasal lexical items used in the test

	Phrasal lexical items
1	Avoid NP like the plague
2	Act the goat
3	Give NP the creeps
4	Drive NP to drink
5	Join/enter the fray
6	Toe the company line
7	Goad/spur NP into action
8	Worship the ground NP walks on
9	Do things by halves
10	Let NP into a secret
11	Make tracks (for)
12	Pluck/summon up courage

13	Seal NP's fate
14	Take a fancy to NP
15	Keep a straight face
16	Spare no expense
17	Wring NP's neck
18	Tighten NP's belt
19	Wipe NP off the map
20	Scrape the bottom of the barrel

The ranges of verbs used by respondents in the cloze items are listed in Table 6.2 – 6.21.

Native speaker consultants' mean rankings (ranging from 1-6) are in the third column.

Table 6.2 List of responses for Case 1 - *Shannon normally _____ these kind of events like the plague, but her pushy new flatmate had convinced her a night out might be in order, especially considering she'd only been at the firm for a few weeks.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	Felt	1.3
2	Refused	2
3	Had	1
4	Disliked	2.4
5	Hated	2.6
6	Dreaded	4.2
7	Attended	1.2

8	Joined	1.1
9	Treated	2.3
10	Thought	1.3

Table 6.3 List of responses for Case 2 - *On closer inspection, she realised she knew a few others at the table – Jenny, who had a tendency to _____ the goat at inter-departmental health and safety meetings, pulling faces and telling stupid jokes*;..

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	be	2.9
2	feed	1.5
3	play	5.7
4	examine	1.3
5	lead	1.3
6	pull	1.8
7	take	1
8	have	1

Table 6.4 List of responses for Case 3 - ..*Annabel, who always looks like a startled deer when you ask her anything that isn't work-related, her face going blotchy at the prospect of real conversation; and Jonno, who _____ every woman on the floor the creeps with his fake smile and lame innuendo.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**

1	made	1.1
2	pretended	1
3	drove	1.1
4	took	1
5	was	1
6	flirted	1.1
7	dated	1
8	thought	1.1
9	posed	1
10	became	1
11	scared	1.3

Table 6.5 List of responses for Case 4 - *Shannon nodded, smiled wanly and mumbled something polite. "Blimey! If that's his friendly party manner then he'll _____ the lot of us to drink!"*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	invite	1.1
2	drink	1
3	make	1.4
4	be	1.1
5	persuade	2.3
6	get	2.3
7	keep	1
8	bring	1.7
9	need	1.4

10	push	3.6
11	treat	1

Table 6.6 List of responses for Case 5 - *"Shall we _____ the fray?" he half-shouted.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	perform	1
2	sing	1
3	dance	1
4	hit	2.6
5	get	1.2
6	do	1.3
7	head	1.2
8	start	1.8
9	call	1
10	have	1.4

Table 6.7 List of responses for Case 6 - *She's fanatical about ___ the company line, though.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	promoting	5.4

2	business	1
3	all	2.5
4	following	4
5	to	1.2
6	endorsing	5
7	planning	1.6
8	making	1.5
9	knowing	2.2
10	managing	2.6
11	lending	1.2
12	chatting	1
13	storing	1.1

Table 6.8 List of responses for Case 7 - *And he's the assistant manager out at the warehouse, always trying to _____ the others into action.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	string	1.2
2	pick	1.1
3	maid	1
4	get	2.7
5	put	2.4
6	force	4.7
7	order	3.9
8	bring	2.4
9	gear	2.4

10	persuade	3.3
11	talk	3.5
12	involve	1.4
13	approach	1.1

Table 6.9 List of responses for Case 8 - *"Shame though, seeing as the 'team' don't exactly _____ the ground he walks on."*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	walk	1.1
2	bring	1
3	match	1
4	share	1.7
5	own	1
6	know	1
7	loose	1
8	reflect	1.2
9	grab	1
10	reveals	1
11	fit	1.2
12	like	1.9
13	go	1
14	present	1
15	follow	1.5
16	have	1
17	suit	1

Table 6.10 List of responses for Case 9 - *"You don't _____ things by halves, do you!?", Peter laughed, "change jobs and change countries for a break!"*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	start	1.2
2	divide	1.2
3	think	1
4	split	1.2
5	lost	1
6	break	1.1
7	take	1.9

Table 6.11 List of responses for Case 10 - *"I'll _____you into a secret" she smirked. "I wasn't exactly mad on staying in Australia. I just waited 'til after I'd had more than enough of my job to make the decision."*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	treat	1.6
2	get	1.2
3	bring	1.4
4	tell	1.5
5	keep	1
6	break	1

Table 6.12 List of responses for Case 11 - *So I handed in my notice and _____ tracks for the safety of inner city England.*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	life	1
2	kept	1.3
3	found	1.4
4	had	1.4
5	off	1
6	went	1.2
7	ran	1.6
8	got	1.4
9	my	1

Table 6.13 List of responses for Case 12 - *"...Still, I don't think I'd have ever _____ up the courage to move over there in the first place. Manchester's about the most exotic place I've ever lived."*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1.	beefed	2.7
2.	sucked	2.5
3.	got	5.5
4.	put	1.4
5.	run	1.2

6.	given	1
7.	built	4.7
8.	mustered	4.9
9.	stood	1
10.	brought	1.5

Table 6.14 List of responses for Case 13 -*"Yeah, I think the snake in my bed after a bad day at work _____ my fate -- I booked my flight home the next day!"*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	with	1
2	is/was	2.4
3	changed	3
4	twisted	1.7
5	decided	5
6	revealed	2.2
7	predicted	2.6
8	cursed	1.1
9	as	1
10	presented	1.4
11	on	1

Table 6.15 List of responses for Case 14 - *"Shall we go back to the table for a bit?" Shannon nodded readily, following him back to the table. She knew she was beginning to _____ a fancy to the guy, and a little group conversation might help her from going overboard.*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	give	1.3
2	notice	1.1
3	full	1
4	have	1.6
5	catch up	1
6	develop	4.1
7	grow	2
8	be	1.1
9	become	1.3
10	make	1.3
11	start	1.4
12	look	1

Table 6.16 List of responses for Case 15 - *Nice dance, was it?" Her supervisor's comment turned her face bright red, and when she looked up Tom was desperately trying to _____ a straight face over the tactlessness of their boss.*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	make	2.1
2	avoid	1.1
3	put	1.7
4	give	1.3
5	pull	1.8

6	show	2.4
7	look	1.2
8	find	1.1

Table 6.17 List of responses for Case 16 -*"They've _____ no expense, have they?" The supervisor picked up the '99p shop' decorations in the centrepiece.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	got	1.1
2	been	1
3	at	1
4	made	1.3
5	used	1.4
6	tried	1
7	left	1.5
8	spent	1.1

Table 6.18 List of responses for Case 17 - *Everybody at the table had a good laugh, but Shannon knew Jake really wanted to _____ Peter's neck.*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	bite	2
2	pull	1.6
3	hold	1.6

4	strangle	2.8
5	get	1.1
6	snap	3.6
7	check	1.1
8	cling on	1.4
9	have	1
10	choke	2.6
11	break	4.8

Table 6.19 List of responses for Case 18 - *"Seriously, though, aren't they trying to _____ their belts? You know, less spending on frivolous parties and more on real staff benefits."*

	Verb/Responses	**Native speaker consultants mean rankings (ranging from 1-6)**
1	wear	1.3
2	tie	1.5
3	buckle	1.8
4	put	1.2
5	loosen	1.1
6	pull	1.5
7	fasten	2.1
8	fit	1.3

Table 6.20 List of responses for Case 19 - *This year, however, we thought we'd _____ last year's Christmas party off the map!*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	throw	1.6
2	celebrate	1.2
3	deal	1
4	take	1.7
5	let	1
6	move	1.6
7	blast	5
8	kick	3.5
9	blow	4.8
10	remove	1.7
11	have	1
12	put	1.5
13	our	1

Table 6.21 List of responses for Case 20 - *There were a few moans about the company _____ the bottom of the barrel with a bus trip before two airline coaches drew up and opened their doors.*

	Verb/Responses	Native speaker consultants mean rankings (ranging from 1-6)
1	policy down	1
2	at	1.6

3	wait	1
4	giving	1.1
5	party	1
6	sinking	1.4
7	from	1.1
8	related to	1

6.4 Analysis

The first hypothesis to be tested was that individuals who have higher scores of idiomatic responses will also have higher plausibility scores for those responses where they do not cloze on the idiomatic alternative. This would require mean and standard deviation analysis. A correlation analysis was also performed in order to observe the relationship between the two variables, i.e. number of idiomatic responses and mean plausibility rank of other responses. The mean was calculated based on individuals' responses. All individuals' responses on each of the twenty cloze items are listed in tables 6.22 - 6.41.

The second hypothesis put forward in the previous section was that increased exposure leads to the acquisition of more lower frequency verbs and thus to a higher rate of their use in the cloze gaps. The results may allow us to infer the verb selection made by the students.

Table 6.22 Student 1 (T1) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible	Standard Deviation

			(Mean)	
1. avoid	avoid	√		
2. act	be		2.9	
3. give	gives	√		
4. drive	invite		1.1	
5. join/enter	perform		1.0	
6. toe	promoting		5.4	
7. galvanise/ goad/ spur	string		1.2	
8. worship	walk		1.1	
9. do	start		1.2	
10. let	treat		1.6	
11. make	life		1.0	
12. pluck/summon	beef		2.7	
13. seal	will seal	√		
14. take	give		1.3	
15. keep	keep	√		
16. spare	got		1.1	
17. wring	bite		2.0	
18. tighten	tighten	√		
19. wipe	throw		1.6	
20. scrape	policy down		1.0	
Total		5	1.75	1.18

Table 6.23 Student 2 (T2) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation

1. avoid	feels		1.3	
2. act	feed		1.5	
3. give	makes		1.1	
4. drive	invite		1.1	
5. join/enter	sing		1.0	
6. toe	-		-	
7. galvanise/ goad/ spur	pick		1.1	
8. worship	brings		1.0	
9. do	divide		1.2	
10. let	let	√		
11. make	keep		1.3	
12. pluck/summon	suck		2.5	
13. seal	with		1.0	
14. take	notice		1.1	
15. keep	make		2.1	
16. spare	got		1.1	
17. wring	break		4.8	
18. tighten	tighten	√		
19. wipe	celebrate		1.2	
20. scrape	at		1.6	
Total		2	1.53	.94

Table 6.24 Student 3 (T3) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation

1. avoid	refuse		2.0	
2. act	be		2.9	
3. give	pretend		1.0	
4. drive	drunk		1.0	
5. join/enter	dance		1.0	
6. toe	business		1.0	
7. galvanise/ goad/ spur	maid		1.0	
8. worship	match		1.0	
9. do	think		1.0	
10. let	get		1.2	
11. make	-		-	
12. pluck/ summon	get		5.5	
13. seal	is		2.4	
14. take	full		1.0	
15. keep	avoid		1.1	
16. spare	been		1.0	
17. wring	break		4.8	
18. tighten	wear		1.3	
19. wipe	deal		1.0	
20.scrape	wait		1.0	
Total		0	1.70	1.34

Table 6.25 Student 4 (T4) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	have		1.0	

2. act	be		2.9	
3. give	drives		1.1	
4. drive	make		1.4	
5. join/enter	hit		2.6	
6. toe	all		1.3	
7. galvanise/ goad/ spur	get		2.7	
8. worship	share		1.7	
9. do	split		1.2	
10. let	bring		1.4	
11. make	find		1.4	
12. pluck/summon	put		1.4	
13. seal	changed		3.0	
14. take	have		1.6	
15. keep	put		1.7	
16. spare	at		1.0	
17. wring	break		4.8	
18. tighten	tighten	√		
19. wipe	take		1.7	
20. scrape	at		1.6	
Total		1	1.87	.95

Table 6.26 Student 5 (T5) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoids	√		
2. act	play		5.7	

3. give	took		1.0	
4. drive	drive	√		
5. join/enter	join	√		
6. toe	following		4.0	
7. galvanise/ goad/ spur	put		2.4	
8. worship	own		1.0	
9. do	do	√		
10. let	let	√		
11. make	made	√		
12. pluck/summon	plucked	√		
13. seal	twisted		1.7	
14. take	take	√		
15. keep	give		1.3	
16. spare	made		1.3	
17. wring	break		4.8	
18. tighten	tighten	√		
19. wipe	wipe	√		
20. scrape	scraped	√		
Total		11	2.58	1.80

Table 6.27Student 6 (T6) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	dislike		2.4	
2. act	be		2.9	
3. give	is		1.0	

4. drive	be		1.1	
5. join/enter	get		1.2	
6. toe	to		1.2	
7. galvanise/ goad/ spur	force		4.7	
8. worship	know		1.0	
9. do	lost		1.0	
10. let	get		1.2	
11. make	had		1.4	
12. pluck/summon	get		5.5	
13. seal	is		2.4	
14. take	have		1.6	
15. keep	make		2.1	
16. spare	used		1.4	
17. wring	pull		1.6	
18. tighten	tie		1.5	
19. wipe	a		1.0	
20. scrape	at		1.6	
Total		0	1.89	1.23

Table 6.28 Student 7 (T7) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	disliked		2.4	
2. act	examine		1.3	
3. give	flirt		1.1	
4. drive	persuade		2.3	

5. join/enter	do		1.3	
6. toe	-		-	
7. galvanise/ goad/ spur	order		3.9	
8. worship	loosed		1.0	
9. do	do	√		
10. let	tell		1.5	
11. make	keep		1.3	
12. pluck/ summon	give		1.0	
13. seal	-		-	
14. take	catch up		1.0	
15. keep	make		2.1	
16. spare	-		-	
17. wring	hold		1.6	
18. tighten	-		-	
19. wipe	celebrated		1.2	
20. scrape	at		1.6	
Total		1	1.64	.78

Table 6.29 Student 8 (T8) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoid	√		
2. act	be		2.9	
3. give	gives	√		
4. drive	get		2.3	
5. join/enter	join	√		

6. toe	promoting		5.4	
7. galvanise/ goad/ spur	put		2.4	
8. worship	reflect		1.2	
9. do	do	√		
10. let	let	√		
11. make	off		1.0	
12. pluck/summon	build		4.7	
13. seal	decide		5.0	
14. take	develop		4.1	
15. keep	put		1.7	
16. spare	got		1.1	
17. wring	break		4.8	
18. tighten	buckle		1.8	
19. wipe	let		1.0	
20. scrape	giving		1.1	
Total		5	2.7	1.65

Table 6.30 Student 9 (T9) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoid	√		
2. act	be		2.9	
3. give	gives	√		
4. drive	make		1.4	
5. join/enter	join	√		
6. toe	endorsing		5.0	

7. galvanise/ goad/ spur	put		2.4	
8. worship	reflect		1.2	
9. do	do	√		
10. let	get		1.2	
11. make	off		1.0	
12. pluck/summon	put		1.4	
13. seal	decided		5.0	
14. take	grow		2.0	
15. keep	make		2.1	
16. spare	got		1.1	
17. wring	strangle		2.8	
18. tighten	buckle		1.8	
19. wipe	let		1.0	
20. scrape	giving		1.1	
Total		4	2.09	1.30

Table 6.31 Student 10 (T10) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviations
1. avoid	hate		2.6	
2. act	lead		1.3	
3. give	gave	√		
4. drive	keep		1.0	
5. join/enter	do		1.3	
6. toe	planning		1.6	

7. galvanise/ goad/ spur	bring		2.4	
8. worship	grab		1.0	
9. do	do	√		
10. let	keep		1.0	
11. make	kept		1.3	
12. pluck/summon	put		1.4	
13. seal	reveal		2.2	
14. take	be		1.1	
15. keep	put		1.7	
16. spare	tried		1.0	
17. wring	get		1.1	
18. tighten	tighten	√		
19. wipe	move		1.6	
20. scrape	scraped		1.6	
Total		3	1.48	.50

Table 6.32 Student 11 (T11) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoid	√		
2. act	be		2.9	
3. give	makes		1.1	
4. drive	get		2.3	
5. join/enter	join	√		
6. toe	promoting		5.4	

7.galvanise/ goad/ spur	put		2.4	
8. worship	reveals		1.0	
9. do	break		1.1	
10. let	bring		1.4	
11. make	off		1.0	
12. pluck/summon	muster		4.9	
13. seal	decided		5.0	
14. take	become		1.3	
15. keep	make		2.1	
16. spare	got		1.1	
17. wring	strangle		2.8	
18. tighten	put		1.2	
19. wipe	let		1.0	
20. scrape	giving		1.1	
Total		2	2.17	1.50

Table 6.33 Student 12 (T12) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	dreaded		4.2	
2. act	pull		1.8	
3. give	gave	√		
4. drive	bring		1.7	
5. join/enter	join	√		
6. toe	making		1.5	
7. galvanise/ goad/	gear		2.4	

spur				
8. worship	fit		1.2	
9. do	do	√		
10. let	break		1.0	
11. make	made	√		
12. pluck/summon	put		1.4	
13. seal	sealed	√		
14. take	make		1.3	
15. keep	pull		1.8	
16. spare	leaving		1.5	
17. wring	snap		3.6	
18. tighten	tighten	√		
19. wipe	blast		5.0	
20. scrape	scraping	√		
Total		7	2.18	1.27

Table 6.34 Student 13 (T13) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoid	√		
2. act	be		2.9	
3. give	gave	√		
4. drive	need		1.4	
5. join/enter	do		1.3	
6. toe	promoting		5.4	
7. galvanise/ goad/ spur	persuade		3.3	

8. worship	fit		1.2	
9. do	do	√		
10. let	tell		1.5	
11. make	went		1.2	
12. pluck/summon	give		1.0	
13. seal	predict		2.6	
14. take	have		1.6	
15. keep	make		2.1	
16. spare	spent		1.1	
17. wring	get		1.1	
18. tighten	tie		1.5	
19. wipe	kicked		3.5	
20. scrape	party		1.0	
Total		3	1.98	1.20

Table 6.35 Student 14 (T14) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoid	√		
2. act	pull		1.8	
3. give	gives	√		
4. drive	push		3.6	
5. join/enter	head		1.2	
6. toe	knowing		2.2	
7. galvanise/ goad/ spur	talk		3.5	
8. worship	walk		1.1	

9. do	do	√		
10. let	let	√		
11. make	run		1.6	
12. pluck/ summon	build		4.7	
13. seal	curse		1.1	
14. take	have		1.6	
15. keep	put		1.7	
16. spare	spare	√		
17. wring	bite		2.0	
18. tighten	loosen		1.1	
19. wipe	blow		4.8	
20. scrape	sank		1.4	
Total		5	2.23	1.28

Table 6.36 Student 15 (T15) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoids	√		
2. act	be		2.9	
3. give	date		1.0	
4. drive	treat		1.0	
5. join/enter	start		1.8	
6. toe	promoting		5.4	
7. galvanise/ goad/ spur	force		4.7	
8. worship	like		1.9	

9. do	divide		1.2	
10. let	keep		1.0	
11. make	kept		1.3	
12. pluck/ summon	given		1.0	
13. seal	was		2.4	
14. take	be		1.1	
15. keep	make		2.1	
16. spare	spent		1.1	
17. wring	break		4.8	
18. tighten	tight	√		
19. wipe	remove		1.7	
20. scrape	from		1.1	
Total		2	2.08	1.44

Table 6.37 Student 16 (T16) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	attend		1.2	
2. act	be		2.9	
3. give	thinks		1.1	
4. drive	get		2.3	
5. join/enter	call		1.0	
6. toe	managing		2.6	
7. galvanise/ goad/ spur	involve		1.4	
8. worship	go		1.0	
9. do	take		1.9	

10. let	let	√		
11. make	get		1.4	
12. pluck/ summon	muster		4.9	
13. seal	as		1.0	
14. take	have		1.6	
15. keep	show		2.4	
16. spare	got		1.1	
17. wring	bite		2.0	
18. tighten	pull		1.5	
19. wipe	have		1.0	
20. scrape	from		1.1	
Total		1	1.76	.97

Table 6.38 Student 17 (T17) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	join		1.1	
2. act	take		1.0	
3. give	pose		1.0	
4. drive	invite		1.1	
5. join/enter	start		1.8	
6. toe	lending		1.2	
7. galvanise/ goad/ spur	force		4.7	
8. worship	present		1.0	

9. do	divide		1.2	
10. let	let	√		
11. make	keep		1.3	
12. pluck/summon	stand		1.0	
13. seal	present		1.4	
14. take	become		1.3	
15. keep	look		1.2	
16. spare	spend		1.1	
17. wring	check		1.1	
18. tighten	fasten		2.1	
19. wipe	celebrate		1.2	
20. scrape	from		1.1	
Total		1	1.42	.84

Table 6.39 Student 18 (T18) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	avoided	√		
2. act	be		2.9	
3. give	flirted		1.1	
4. drive	make		1.4	
5. join/enter	start		1.8	
6. toe	chatting		1.0	
7. galvanise/ goad/ spur	bring		2.4	
8. worship	follow		1.5	
9. do	break		1.1	

10. let	keep		1.0	
11. make	kept		1.3	
12. pluck/ summon	brought		1.5	
13. seal	changed		3.0	
14. take	start		1.3	
15. keep	show		2.4	
16. spare	spent		1.1	
17. wring	cling on		1.4	
18. tighten	tight	√		
19. wipe	had		1.0	
20. scrape	at		1.6	
Total		2	1.6	.65

Table 6.40 Student 19 (T19) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	treated		2.3	
2. act	have		1.0	
3. give	became		1.0	
4. drive	invite		1.1	
5. join/enter	have		1.4	
6. toe	-		-	
7. galvanise/ goad/ spur	approach		1.1	
8.worship	have		1.0	
9. do	divide		1.2	

10. let	hire		1.0	
11. make	my		1.0	
12. pluck/summon	put		1.4	
13. seal	on		1.0	
14. take	look		1.0	
15. keep	find		1.1	
16. spare	-		-	
17. wring	have		1.0	
18. tighten	tight	√		
19. wipe	our		1.0	
20. scrape	related to		1.0	
Total		1	1.15	.32

Table 6.41 Student 20 (T20) responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Semantically plausible (Mean)	Standard Deviation
1. avoid	thinks		1.3	
2. act	play		5.7	
3. give	scares		1.3	
4. drive	make		1.4	
5. join/enter	hit		2.6	
6. toe	storing		1.1	
7. galvanise/ goad/ spur	get		2.7	
8. worship	suit		1.0	
9. do	take		1.9	
10. let	let	√		

11. make	keep		1.0	
12. pluck/summon	run		1.6	
13. seal	changed		3.0	
14. take	start		1.4	
15. keep	show		2.4	
16. spare	spent		1.1	
17. wring	choke		2.6	
18. tighten	fit		1.3	
19. wipe	put		1.5	
20. scrape	at		1.6	
Total		1	1.92	1.11

6.5 Results and discussion

6.5.1 Individuals who have more idiomatic responses will also have higher plausibility scores for other responses.

The results show that the semantically plausible means for all 20 students was below 3. The highest mean achieved was 2.7. The results presented in Table 6.42 show means and standard deviations for all 20 respondents whose test results were recoded.

Table 6.42 Percentage of idiomatic responses, semantic plausibility mean and standard deviation of individuals taking the test

Respondent	**Idiomatic (%)**	**Idiomatic (20total)**	**Semantically plausible Mean**	**SD**
Student 1	25%	5	1.75	1.18
Student 2	10%	2	1.53	0.94
Student 3	0%	0	1.70	1.39

Student 4	5%	1	1.87	0.95
Student 5	55%	11	2.58	1.80
Student 6	0%	0	1.90	1.23
Student 7	5%	1	1.64	0.78
Student 8	25%	5	2.7	1.66
Student 9	20%	4	2.09	1.30
Student 10	15%	3	1.48	0.50
Student 11	10%	2	2.17	1.50
Student 12	35%	7	2.18	1.27
Student 13	15%	3	1.98	1.20
Student 14	25%	5	2.23	1.28
Student 15	10%	2	2.08	1.44
Student 16	5%	1	1.76	0.97
Student 17	5%	1	1.42	0.84
Student 18	10%	2	1.6	0.65
Student 19	5%	1	1.15	0.32
Student 20	5%	1	1.92	1.11

Percentages of correct idiomatic scores were categorized as high (>50%), average (15-25%) or low (0-5%). Only one respondent had a 'high' score, with 55% correct. This student's mean plausibility score was 2.58. About eight respondents were categorized as having low scores, with the rest categorized as average. However, the semantic plausibility metric seems to reveal that the respondents have different mean scores. These results led to performance of a correlation analysis. The correlation analysis was performed in order to observe the relationship between the two variables, i.e. number of idiomatic responses and mean semantic plausibility of non-idiomatic answers responses.

The Pearson correlation coefficients (r) take on values from -1 to +1. The motivation was to describe the strength and direction of the relationship between the variables. The following tables show the descriptive statistics of mean and standard deviation for the responses of 20 respondents for the study, and the summary of correlation results.

Table 6.43 Means and standard deviations for responses of 20 respondents

	Respondents (N= 20)	
	Mean	Standard deviation
Idiomatic responses	2.85	2.70
Mean of non-idiomatic responses	1.89	.38

Table 6.44 Summary of correlation results for idiomatic responses and mean non-idiomatic responses for 20 respondents

N=20	
Pearson correlation of idiomatic responses and mean of non-idiomatic responses	r = .652
p	.0009

The results revealed that the relationship between the number of idiomatic responses and the mean plausibility of the non-idiomatic responses was strong, with a positive relationship observed between the variables, r = .652, and, N = 20, p= .0009. This

suggests that individuals who have higher numbers of idiomatic responses will also have higher plausibility scores on their non-idiomatic responses. This corresponds to the direction of the predicted correlation.

6.5.2 Case studies - samples of individuals' sets of responses

In this section three students' answers are presented and discussed in detail. Anna's and two other respondents' non-idiomatic responses were analyzed using the mean results. Anna and two other respondents' answers were examples of high, middle and low idiomatic responses. Their responses will be analyzed and compared to each other. The following table shows Anna's responses for both the idiomatic and non-idiomatic answers.

Table 6.45 Student 5 (T5) –Anna's responses

Idiomatic verb	Student's answer/verb	Idiomatic target	Non-idiomatic response (Mean)	Standard Deviation
1.avoid	avoids	√		
2.act	play		5.7	
3.give	took		1.0	
4.drive	drive	√		
5.join/enter	join	√		
6.toe	following		4.0	
7.galvanise/goad/spur	put		2.4	
8.worship	own		1.0	

9.do	do	√		
10.let	let	√		
11.make	made	√		
12.pluck/summon	plucked	√		
13.seal	twisted		1.7	
14.take	take	√		
15.keep	give		1.3	
16.spare	made		1.3	
17.wring	break		4.8	
18.tighten	tighten	√		
19.wipe	wipe	√		
20.scrape	scraped	√		
Total		11	2.58	1.80

Anna's score for the idiomatic responses was 55% or equal to 11 out of 20 idiomatic responses, and was the highest score of all 20 respondents. Her mean score was 2.58 for the non-idiomatic responses. This was the second highest among all respondents and this has led to analyzing all her semantically plausible responses. The following list shows her semantic plausibility verb selections for filling up the gaps.

Anna filled in the gaps of the cloze test with 9 non-idiomatic answers, listed below:

1. ACT the goat- PLAY
2. GIVE NP the creeps- TOOK
3. TOE the company line- FOLLOWING
4. GALVANISE/ GOAD/ SPUR – PUT
5. WORSHIP the ground NP walks on- OWN
6. SEAL NP's fate- TWISTED
7. KEEP a straight face- GIVE
8. SPARE no expense- MADE
9. WRING NP's neck –BREAK

The non-idiomatic verbs were checked using a frequency rank in Kilgarriff's lemmatized BNC frequency list. There were 6,318 words in the lemmatized frequency list including 1,281 verbs. I have grouped the verbs into bands with ten verbs per band, and Anna's responses were ranked based on these bands. The bands were restricted to the first 12 bands on the presumption that the first 120 verbs (10%) are at the high end of the frequency spectrum. The following table shows the appearance of the verbs based on the bands. Their frequency and sort order in Kilgarriff's lemmatized list are also listed in Table 6.46. An example of a verb band is given in Table 6.47.

Table 6.46 Band, sort order and frequency of the plausible verb responses

Verb	Band/12	Sort order	Frequency
PLAY	5	245	38,058
TOOK	1	54	179,220
FOLLOWING	4	203	46,145
PUT	2	125	69,978
OWN	>12	1,536	6236
TWISTED	>12	3,480	2004
GIVE	2	76	131,417
MADE	1	46	217,268
BREAK	11	532	19,512

Table 6.47 Band 1 of the first 10 verbs extracted from Kilgarriff's lemmatized BNC frequency list

No	Verb	Rank no
1	be	2
2	have	8
3	do	18
4	say	34
5	go	40
6	get	44
7	make	46
8	see	51
9	know	52
10	take	54

The results show that 7 plausible answers provided by Anna were among the 120 verbs in the first 12 bands. Only 2 verbs, i.e. *own* and *twisted*, were below those bands in frequency. The results reveal that 77.8% of the plausible verb choices made by Anna were highly frequent and could be categorized within the highest verb frequency category. This suggests that the second hypothesis, that the verb choice made for the non-idiomatic answers would be at the high end of the frequency spectrum, was supported.

A second example was from respondent number 8 (refer to Table 6.29) and a third from respondent number 6 (refer Table 6.27). Respondent 8 was considered to represent the average score group, while Respondent 6 was representing the low score group, i.e. scored 0-1 (or 0% to 5%) of the idiomatic responses. Student 8, referred to as Reyna,

scored 25% of the idiomatic responses and the mean score for semantic plausibility was 2.7, which was slightly higher than Anna' mean score, and the highest among other respondents. Table 6.46 shows scores of three respondents representing the highest, middle and low score group.

Table 6.48 Scores of three respondents representing the highest, middle and low score group.

Respondents	Idiomatic target	Non-idiomatic responses (Mean)
Anna (Respondent 5)	55%	2.58
Reyna (Respondent 8)	25%	2.7
Eusoff (Respondent 6)	0%	1.89

The results reveal that even though Anna's score was 55% on the idiomatic responses, her mean score on the non-idiomatic responses was 2.58. Reyna's score shows that even though a learner fails to get a good score for the idiomatic targets, it is still possible for learners to reach 'the level of semantically plausible responses' as these were considered acceptable responses and 'close' to the idiomatic target by the native speakers of English. Eusoff's idiomatic target score was 0% and his mean score on the non-idiomatic responses was 1.89 with only two most plausible responses used in the test. Yet these two verbs received a very high plausibility rate. While the head verbs *galvanise/goad/spur* and *pluck/summon,* are low frequency verbs, the high plausibility rating for the respondents' answers were surprisingly good. The selections of the head verbs of force and get as alternatives were seen to be plausible and acceptable by native speakers of

English. And in Reyna's case there were five highly rated semantically plausible verbs as listed in Table 6.50.

Table 6.49 Eusoff's most semantically plausible responses

Idiomatic verbs	Student's responses	Mean score of semantically plausible answers
Galvanise / goad / spur	force	4.7
Pluck / summon	get	5.5

If we refer back to Table 6.29, Reyna had quite a number of high mean scores of semantically plausible responses. There were about five of them, all listed below with the mean scores:

Table 6.50 Reyna's most semantically plausible responses

Idiomatic verbs	Student's responses	Mean score of semantically plausible answers
toe	promoting	5.4
pluck/ summon	build	4.7
seal	decide	5.0
take	develop	4.1
wring	break	4.8

6.5.3 Discussion

The results presented above support the hypotheses that:

1. Individuals who have the higher scores of idiomatic responses will also have higher scores for the plausibility of their non-idiomatic responses.

2. Given the presumption of less exposure to the target language, more of the non-idiomatic responses will be at high end of the frequency spectrum

The data presented in the first case also supported Dornyei and Skehan's (2003) as well as Sawyer and Ranta's (2001) notion of the impact of individual difference on language learning in general. It will indirectly influence the vocabulary acquisition of each individual. There was evidence of a reasonably predictable knowledge of individual lexical items, shown in Table 6.42. What is more, the *idiomatic* column presented the respondents' knowledge of collocations and the mean semantic plausibility of non-idiomatic responses derived individually to reflect individuals' lexical knowledge. This observation is closely related to language processing. So, in this context, when the respondents were faced with a lexical retrieval task, they applied two possible strategies in retrieving specific vocabulary: either retrieving whole phrases or single words. In general most of their mean scores were below 3.0, which may be considered quite low, yet when judging the semantic plausibility of their responses there was evidence that general vocabulary acquisition had taken place although their range of lexical items was quite limited. That may explain why respondents opted for non-idiomatic responses. Furthermore, most research indicates that non-native speakers often have relatively weak mastery over formulaic language, in some cases resulting in misuse (Howarth, 1998), under-use (Dagut & Laufer, 1985), or even overuse (Granger, 1998). These findings are supported in the present research, since students produced inappropriate and awkward

choices to fill the gaps where they did not know the idiomatic verb filler. This is reflected in the analysis where no respondent scored over 3 in their mean score in the case of non-idiomatic selection. Schmitt, Grandage and Adolphs (2004) suggest that there are certain strategies adopted by non-native speakers when they have difficulty reproducing the dictated text. They state that, "with limited memory capacity in their L2 and language competence which inevitably had some limitations, the non-native participants seemed to 'latch onto' key content words and then try to reproduce the dictation language around them. They did not seem to have the recurrent clusters available as formulaic sequences, and so tried *to generate a sensible reconstruction* based on these key words" (pp. 140)

The strategy being adopted resulted in quite a number of learners producing *incorrect* or *disfluent* text where their reproduced sequences were quite different from the target. A similar trend can also be observed in the present study. A good example is Reyna's case where she was able to supply five most the plausible verbs for the test. Even though her score for the idiomatic responses was quite low, her vocabulary knowledge resulted in a higher mean score for plausibility. It shows that though non-native speakers might not be able to supply the most idiomatic responses, they may still provide sensible responses. This might result from having received adequate exposure to the target language but not sufficient to provide idiomatic responses.

A non-native speaker like Anna does not have the advantages of native speakers whose number of fixed expressions stored in mental lexicon is vast (Jackendoff, 1995; Melčuk, 1995: 169), although it is hard to have an accurate estimation of the extent of the

formulaic language stored in the mental lexicon (Kuiper, Columbus, & Schmitt, 2009). Kuiper, Columbus & Schmitt, (2009) further argue that there is a possibility that there are a larger number of phrasal lexical items than single word items in a native speaker vocabulary. So, if learners like Anna, Reyna and Eusoff do not have enough PLIs in their mental lexicon, native-like competency is hindered, requiring them to opt for other strategies when having language difficulties.

Another related strategy which might be applied by the participants is *guessing from context* which is commonly used for learning individual words (Nation, 1990, 2001). This can be observed from the most non-idiomatic responses provided by the learners in this study. A good example to refer to is Table 6.40, where this respondent's answers were believed to be based on wild guesses. None of this respondent's means scores were more than 2.3.

Native speakers have the advantage in processing speech in that they have access to expressions stored as ready-made wholes. However, non-native speakers do not share the same processing advantage. Schmitt, Grandage and Adolphs (2004) found that recurrent clusters had to be reconstructed rather than being repeated by rote from memory. In that research, the task required language learners to repeat the dictated bursts exactly and the learners were expected to draw upon any of the target clusters they had stored in memory. It was assumed that they were stored as wholes in memory and would be repeated fully intact, without hesitation and with a normal stress profile. It was found that many of those recurrent sequences were not repeated in such a manner, or even produced at all by the

speakers. The evidence from this research is that these recurrent sequences may not in fact be stored as formulaic sequences in the minds of these participants. Thus the guessing game strategy was applied.

The above findings were closely related to the second hypothesis, namely that less exposure leads to the choice of high frequency verbs in filling the cloze gaps. In this case verbs like *took (take), made (make), give* and *put* were chosen as alternative head verbs, as shown in Table 6.46. These verbs were ranked among the first 20 verbs of Kilgarriff's lemmatized list. So, if the strategy was to guess from the context (Nation, 1990, 2001), the guessed verbs were from the highest frequencies. But, if more exposure was received the prediction is that more verbs beyond the 12th band will be chosen by a respondent. The verbs like *own* and *twisted* reflect the respondent's exposure to the less frequent lexical items. This suggests that exposure played significant role in the language production and lexical retrieval task involved in the cloze test. I have no evidence to suggest that Anna and the other two respondents are typical (Malaysian) learners and the results should not be generalized to represent other second language learners studying overseas. However, what we have seen here is that this study gives insight into learners' test taking strategies. The results suggest that further investigations appear to be worthwhile for other cohorts and more individuals to gain further insight into the nature of lexical knowledge. Thus, I am proposing that the semantic plausibility metric which is used as a tool for this study can be useful used as a measure of vocabulary acquisition. To be specific, it is a tool measuring learners' test taking strategies. The three students'

profiles reflect a significant finding of lexical knowledge and vocabulary acquisition in general.

6.5.4 Conclusion

The findings in this chapter are significant because they illustrate the types of responses learners tend to come up with and indirectly illustrate the challenge of mastering restricted collocations. What is observed is that many other verbs than the appropriate ones are semantically plausible candidates. However, these are arbitrarily blocked from the standard, exclusive word part partnership. But if so many responses are semantically plausible, would their use cause a strain on communication, to be specific, communication among non-native speakers or Outer Circle users of English? The study by Millar (2011) has documented that the malformed L2 collocations lead to an increased processing burden for native speakers in terms of slower reading speed. So, it does put some strain on native-speakers' processing. However, if we view L2 use from a lingua franca perspective, native-like attainment and selection may possible not necessarily be the aim for L2 development. In this sense, the malformed or infelicitous restricted collocational choices made by L2 learners should in fact be viewed more positively as instances of risk-taking strategy in order to cope communicatively. Among the Outer Circle users of English the strain on processing will not be experienced because the chunks are not entrenched in their mental lexicon. So, the malformed collocations could be a way of making the World English perspective relevant after all. To this end, an interesting question would be raised regarding the need for L2 standard English collocation teaching and learning.

Thus, Kachru's model of Concentric Circles (1985, 1992) has serious implications not just for the way English conceptualized in the world context but also on the teaching and learning of collocations. As this study focuses on the factor of exposure to both Englishes i.e. Outer Circle of English which is represented by English of Malaysia, and Inner Circle English (New Zealand English), this study reinforces that there are challenges of mastering restricted collocations and the use of these sequences. May be within the same speech community standard collocations are less used and expected, but for academic writing and purposes, it has been generally agreed that the appropriate use of these sequences is highly required (Li and Schmitt, 2009; Howarth, 1998).

CHAPTER 7

CONCLUSION, SUGGESTIONS AND TEACHING IMPLICATIONS

7.1 Introduction

This study examined the lexical collocations acquired by Malaysian learners who had exposure to both Malaysian English and New Zealand English. The focus was restricted to Verb-Noun collocations of written English. If we raise a question of what actually facilitates acquisition of formulaic sequences in general, there appear to be no fixed answers. Would the size of the mental lexicon in terms of individual word forms predict the acquisition of PLIs? Would the frequency of the verbs or the frequency of the restricted collocations play a significant role in the acquisition? How does studying abroad affect acquisition? This thesis has attempted to answer some of these questions.

This chapter will outline the overall summary of all findings retrieved from the studies done while completing the whole project. Together with pedagogical aspects, the strengths and limitations of the study are highlighted. Apart from that, this chapter will discuss the implications and applications of the present study based on research findings reported in Chapters 3, 4, 5 and 6. The summary notes will be addressed by several subheadings and a few suggestions for future research are made.

7.2 Corpus approach

The fact is that there are distinctive benefits of using corpus data in language studies. I observe similar effects in this study. This corpus not just provides various ways of approaching the studies from different perspectives, but also provides a valuable resource

as a collection of endonormative data, i.e. local news reports in English. Once the NST corpus is balanced and further developed, it will become useful to linguists, researchers and teachers for doing related tasks in language studies.

7.3 Exposure to the target language collocations

It is widely agreed that exposure to language plays a significant role in the acquisition process (Adolph & Durow, 2004; Krashen, 1982; Ellis, 1994). The influence of exposure was explored from two different perspectives; exposure to the collocations of written Malaysian English, and exposure to standard English i.e. New Zealand English. Both studies looked at the duration of time, year of study and months of learning, so as to measure exposure. Another factor which was highlighted was the age-graded factor in acquisition. Age in this sense is considered as a proxy to exposure.

The results of the analysis of variance or ANOVA in section 4.4.1 of Chapter 4 reveal that the interaction effect between four groups of participants and verb types was statistically significant. Another finding was that there was a statistically significant main effect for verb type, suggesting differences in scores between the 4 different groups of learners among those 202 students. All pairwise tests were statistically significant ($p < .05$) with the exception of the difference between UUM (Malaysian tertiary group) and secondary school. In other words, there were differences of vocabulary acquisition among the school students, university students in Malaysia, Malaysian students who were studying in New Zealand universities and native speakers of New Zealand English.

The ANOVA results in Chapter 5 appear to reveal that exposure of the kind of experienced by international students does not have significant effects on the acquisition of restricted collocations.

When we ask what related factors may contribute to this result the answer would be that the acquisition of restricted collocations is so much associated with learners' active involvement in some of the native social community (Dörnyei, Durow & Zahran, 2004). So in the case of learners in Study 2, these international students may not be actively involved in local social community where English is the medium of instruction. They might find it hard to join such 'host-national networks' as they come from different cultural background. This is closely related to self-confidence which is another important dimension which facilitates the motivational process in multi-ethnic settings (Cléments, 1980).

The study's findings in Chapter 5 do not support the prediction that year of study has an impact on the acquisition of restricted collocations. However, Kuiper, Columbus, & Schmitt (2009), and Escaip (2008) show that acquisition of restricted collocations is age-graded, but over much longer periods. Since the study looks only at three years of exposure to native speaker English, one possible explanation is that this period is too short to have a measurable effect. A second possible explanation is that international students, when there are sufficient numbers of them in an overseas university, form a linguistic and cultural community where their home languages remain the ‘Umgangssprache’, i.e. since our test targeted vernacular vocabulary, it may be that it is

just this vocabulary which is used in day-to-day interaction by international students in their own native languages.

This finding is undesirable for those who promote internationalization of student education on the basis of the view that emersion in a target language community is alone sufficient for students to acquire a more idiomatic phrasal vocabulary. This was mentioned in Kuiper & Kuiper (2003) as a feature of the current tertiary education market under globalization, in which the aim is basically focussing on Asian students gaining locally degrees taught by staff from universities abroad or from English-speaking countries. Kuiper & Kuiper claimed that though these learners have already acquired an endonormative standard of a local English variety, they still hunger for something they perceived as an exonormative standard English from universities abroad.

The findings therefore support other research which indicates that non-native speakers often have relatively weak mastery of formulaic language, resulting in misuse (Howarth, 1998), under-use (Dagut & Laufer, 1985), or overuse (Granger, 1998) of particular items of phrasal vocabulary.

What we can conclude from the findings in Chapter 5 is that exposure *per se* is not in itself a useful concept when we are looking at the acquisition of phrasal vocabulary. Clearly in Chapter 5, the international students who were our respondents had exposure to a native speaker variety of English, but this was not sufficient to alter their acquisition of vernacular phrasal vocabulary over the three years of their study overseas. However,

the frequency of the cloze verb does have an effect, as predicted by Kuiper, Columbus & Schmitt (2009). This is so because frequency is a measure of likely exposure. The more frequent an item is in corpora, the more likely a learner is to be exposed to it.

What is needed is a much more nuanced notion of exposure. The concern is exposure to what varieties and sub-varieties of the target language? In other words we need to look at cloze responses in terms of the target language resource to which learners have access, as discussed in Chapter 6 of the thesis. Learners will normally fill a cloze item with a plausible filler when they understand the cuing text. But it may not be the most native-like filler. So how much exposure do learners have to a particular item of phrasal vocabulary? If acquisition is related to exposure, it must be exposure to particular vocabulary items and many items of phrasal vocabulary have, as we have already suggested, a very low frequency of occurrence even for native speakers (Wray & Perkins, 2000; Wray 2002; Peters, 1983).

The second factor in determining what effect exposure has is just how long is necessary. It is clear that native speakers are exposed to a very large number of dialects and genres of their native language for most of their waking lives, and yet even for them, acquisition of phrasal vocabulary is slow.

Most language learning is not a result of tuition but of exposure. It is an interesting question to ponder how exposure for international students would have to be altered so that language learning outcomes are altered.

If we pose a question regarding the effect of age grade on acquisition, namely Is acquisition age-graded?, both Study 1 and 2 failed to confirm that this is the case. Though Kuiper, Columbus & Schmitt (2009) found that PLI acquisition is age graded, it may not be generalized to the acquisition of PLIs in a 2nd language. What is found in section 5.4.1 of this thesis is that year of studies has no significant effect on the acquisition of restricted collocations. Since there are three groups of students which are divided by Year 1, 2 and 3 of the undergraduate studies, age does not predict acquisition. Maybe what can be concluded is that this period is too short to have an impact. A significant impact might possibly be detected over big gaps of age among the respondents used for the research, i.e. as what was done by Kuiper, Columbus & Schmitt (2009). Using a broader age group of participants for future research is highly recommended in order to see a measurable impact.

Based on findings in Sections 5.4.2 and 5.4.3, months of learning is considered the best predictor of acquisition. The correlation results (see Table 5.4) show that the relationship between months of learning and verb types are strong and positive, where more months of exposure resulted in higher total score performance. In addition to that, there was strong and positive relationship between months of learning and total score, r=.398. However, there was not a strong relationship between either year of study or age with total score performance. The multiple regression results suggest that months of learning contributes to the prediction of acquisition of restricted collocations.

7.4 Frequency effects

The following discussion will address the previously asked question ‘does the frequency of the verbs or the frequency of the restricted collocations play a significant role in acquisition?’ Findings in Chapter 4 and 5 reveal that frequency of restricted collocations may not have a significant effect on the acquisition of such sequences. Both analyses fail to show that frequency of restricted collocations predicts acquisition. The findings in Sections 4.4.2 and 5.4.4 confirm that the frequency of the restricted collocations does not play a significant role in the acquisition of vocabulary.

However, the frequency of head verbs does have an impact on such acquisition. Both section 4.4.1 and 5.4.1 show that there are significant main effect for verb types (or head verbs). In other words, acquisition of restricted collocations is related to head verb frequency. The ANOVA results in both sections revealed that the effect of verb types (or verb frequencies) was observed as participants achieved the highest score on the most frequent category and followed by the other three categories or bands. On a difficulty scale, expressions containing *high light frequency* verbs were considered the easiest ones, followed by expressions with *high frequency* verbs and *medium frequency* verbs, ending with the expressions including *low frequency* verbs, which were considered the hardest ones on the scale. As expected, in chapter 4, the interaction effect between four groups of participants and verb types was statistically significant. There was also a statistically main effect for verb type. These findings suggest that both native speaker and three non-native speakers groups achieved a higher number of correct answers for the restricted collocations with high light frequency verbs than the restricted collocations using low

frequency verbs, with the high and medium frequencies being in the middle. Similar result was observed in Chapter 5.

The results in Chapters 4 and 5 show that the frequency of the clozed head verb of a restricted collocation does make a difference as to whether a restricted collocation is learned. This result follows the findings of Kuiper, Columbus & Schmitt (2009), and Escaip (2008), that the frequency of usage of the head verbs contained in verb plus complement restricted collocation is linked to the acquisition of such sequences. The explanation for this is not as obvious as it may seem. The obvious view is that if a verb has not been acquired then a restricted collocation with it as a verb would also not be acquired. A more subtle interpretation would be that, since restricted collocations are generally not nearly as frequent as the individual words which compose them, it is relatively unlikely that a restricted collocation will be learned at all. It is particularly likely that, where possible, respondents will utilise light verbs given their high frequency and semantic utility in a variety of contexts to provide a semantically plausible reading for the cloze. High frequency light verbs are also head verbs of many more restricted collocations than less frequently occurring verbs, as is shown in the data contained in Kuiper et al. (2003). It is therefore more likely that a guess using an appropriate high frequency light verb will be an idiomatic choice. Careful psycholinguistic enquiry would be needed to decide which of these factors are more likely to have an influence.

7.5 Individual differences

As mentioned earlier, the analytic approach used in Chapter 6 is congruent with the work of Dornyei and Skehan (2003), and Sawyer and Ranta (2001), who have highlighted the finding that individual difference factors have been shown to have significant impacts on language learning in general. Apart from that, Schmitt, Dornyei, Adolphs & Durow (2004) argue that these individual differences might also influence the acquisition of formulaic language. The mean result in Table 6.4 provides support for the above finding. Individual difference is indeed a factor. It is not obviously observed, however, as the mean results for all 20 respondents in their use of non-idiomatic fillers is mostly less than 3, with a few respondents whose scores are more than 2.5, e.g. 2.58 and 2.7. This measure evaluates whether the respondents have general understanding of supplying responses which native speakers would consider semantically plausible enough to fill the cloze gaps. The results show that they still do not have sufficient vocabulary to supply either the idiomatic verb or even a verb that could be considered 'close' enough to be a plausible alternative. This might relate to the non-compositionality of the idioms; their whole meaning cannot be deduced from the meaning of each of their separate lexical constituents. This may prevent learners from analysing some PLIs for their meaning. According to Wray (2002:33), even in extreme cases, a non-compositional sequence is required to be previously learned in order to guarantee the understanding of its meaning.

The above discussion leads to a further question, namely how would native speakers experience cloze test situation when approaching the same test? Where they know the PLIs they will fill the cloze item idiomatically. But are they better at filling a cloze gap

more plausibly than non-native speakers? That remains an area for further research. However, if a Malaysian English cloze test, where the restricted collocations are extracted from the NST corpus, is administered to the native speakers group (as in Chapter 4), would the collocations be easily retrieved as well? If the cuing material was from a variety of English other than the speaker's own, the responses of native speakers to the cloze test would not be as high as it would have been if the cuing text was in their own variety. This shows that the influence of context is significant.

What has been found in section 4.4.1 is that there are differences of mean scores between the native and non-native speakers. Even though the group of native speakers are given a text written in Malaysian English and the restricted collocations are mainly used in Malaysian English writing, native speakers of New Zealand English face no difficulties to retrieve them. This is because the restricted collocations used are also mainly used in standard English. They may have some difficulties understanding the context, but since they are familiar with English restricted collocations the retrieval is absolutely fast. A suggestion for future research would be to apply the same approach that was used for chapter 6, but using native speakers' plausible answers, analysing them with the same approach. This may lead to another discovery in language acquisition.

7.6 Suggestions and teaching implication

If refer back to the conclusion made in Chapter 6 (section 6.5.4) on the implication of World English on the teaching and learning of collocations we might realize that it calls for more strategic and constructive teaching and learning approaches. May be within the

same speech community standard collocations are less used and expected, but for academic writing and purposes, it has been generally agreed that the appropriate use of these sequences is highly required (Li and Schmitt, 2009). We admit that collocations are essential in academic writing and the knowledge of this type of vocabulary is obviously a prerequisite for writing. Thus, the teaching of these sequences is still required, may be at least for the standard English sequences.

Since various types of collocations are stored and processed differently in the mental lexicon, this means that different approaches are needed in teaching and learning these collocations. However, as they are stored as a whole because of their relatively high fixedness then these forms are worth teaching. Nation (2001) suggests that in order to develop fluency all collocations are imperative and learners should encounter these sequences repeatedly. This means that exposure is deemed important in language acquisition (Siyanova and Schmitt, 2007). There are few elements which to which we should draw our attention. There are few guidelines and suggestions made by Boers, Lindstromberg and Eyckmans (2014) regarding the collocational learning in contemporary language pedagogy. They reported that since there are many obstacles to incidental uptake of collocations, the explicit approaches to collocational learning are required. More research related towards testing and comparing the effectiveness of teaching collocations is proposed (Boers and Lindstromberg, 2012). Above all, learners' awareness and attention are vital to collocational leaning, apart from exposure.

I personally, as a teacher and second language user of English, feel that more exposure is needed in the classroom, as well as explicit exposure. Formal and informal exposures are certainly essential to focus more awareness on phrasal items, and restricted collocations specifically. Apart from teachers and learners, curriculum developers and syllabus designers should be aware of the importance of restricted collocations. The right teaching material, including textbooks and corpora, and testing tools for vocabulary acquisition are important to ensure that learners have sufficient exposure. For example, in choosing the right testing tool for vocabulary, NST corpus can be considered as a practical method due to its endonormative standard.

In addition to that, the approach used in Chapter 6 has shown that it is a significantly constructive tool for future research. This approach is novel and has the potential of aiding teachers and researchers to examine the collocations used by learners. It may not only be relevant to learning English but could also be applied to other languages. The use of endonormative data rather than exonormative data for the coding is actually an advantage and is seen as an appropriate way of evaluating learners' plausible responses. According to Kirkpatrick (2007:189), an endonormic model points to the fact that 'a localised version of the language has become socially acceptable'. So, in this sense native speakers of English in New Zealand have coded the plausible answers based on their acceptability. The assessment made by the native speakers is seen as a good recourse and considered as a legitimate source. Another suggestion is that the semantic plausibility metric can be useful as a measure of vocabulary acquisition.

Apart from that, another important finding from the analysis in Chapter 6 is that the way the non-idiomatic responses are recoded leads to a way of testing the level of difficulty of the cloze test used for this study. What is obvious is that some of the semantically plausible responses received much higher plausibility rate from the native speakers. The following table lists the verbs rated as highly plausible rated.

Table 7.1 The highly plausible rated verbs

	Phrasal lexical items	Highly plausible rated verbs	Native speakers consultant mean rankings (ranging from 1-6)
1	avoided NP like the plague	dreaded	4.2
2	act the goat	play	5.7
3	toe the company line	promoting endorsing	5.4 5.0
4	goad/ spur NP into action	force	4.7
5	pluck/ summon up courage	got built mustered	5.5 4.7 4.9
6	seal NP's fate	decided	5.0
7	wring NP's neck	break	4.8
8	wipe NP off the map	blast blow	5 4.8

This has highlighted that apart from fixed idiomatic phrases there are other possible and plausible verbs which can be used as alternatives for learners. So, when it comes to testing learners on these items, these highly rated verbs might become good distractors. They can functionally be good distractors for tests, especially multi choice tests. Thus this would be a help for teachers and language practitioners in designing their vocabulary testing items.

Regarding the international group of students studying in New Zealand, Dörnyei, Durow & Zahran (2004) suggest that integration into the L2 environment and culture can possibly make a difference. Dörnyei, Durow & Zahran (2004) tested 4 'good' formulaic sequence learners and 3 'poor' learners over the course of six months. The results revealed that in order to master formulaic sequences, learners should integrate into the 'host national networks' (p. 104). International students should come into meaningful contact with English speakers outside their academic environment so that they can achieve success in acquiring formulaic sequences.

7.7 Strengths and limitations of the present study

This subchapter will highlight the strengths and limitations of the study as well as addressing some strengths and limitations of the corpus building.

Some of the limitations were related to problems during the data collection period. These problems were found to have an effect on the present study, as the sampling had a slight change because of them. The research was scheduled from 28 February until 18 March 2012, but the research period had to be extended from 15 days to 18 days due to the confirmation of schools' participation in the project. The most obvious problem was commitment and cooperation. Problems occurred when privately owned schools (with English as the medium of instruction) refused to participate in this research project, as they do not feel obligated to do so. They did not want to cooperate or even to negotiate about making an appointment. The school principals kept on giving excuses for not participating. What was worse was that most of them did not reply to emails and some

refused to answer telephone calls. Another problem related to that was the location of schools and time restrictions. I admit that I had problem with locations as they were located in different states in Malaysia. It caused difficulties in adjusting the appointments with the principals as schools' activities were quite tight. Due to these problems only few schools were selected for data collections and none English based school were nominated at all.

In order to investigate the effect of exposure to Malaysian English, it was necessary to obtain data on this variety. To this end, as documented earlier, the NST corpus was developed as part of this study to provide data for this research purpose. It provides further opportunities for research and should be a good reference for other researchers to conduct research in related areas in Malaysian English vocabulary acquisition. It is essential to highlight that NST corpus is not a 'balanced' corpus though it contains more than 19 million words. The major texts are news reports. Hitherto, there are no other existing corpora of Malaysian English. Thus, NST corpus is the only resource for corpus study of Malaysian English. In this case, in the future it is suggested that this corpus be further developed with both spoken and written materials, and that it covers various genres.

This study is limited to investigating the acquisition of restricted collocations of Malaysian English. Research on this English variety is scant, and this research has illustrated a significant literature on this type of acquisition. It would be more interesting to see further research being done of linguistically similar context and scenery.

The studies in this thesis are exploratory and due to the lack of previous studies in the Malaysian context, further research will be needed. In short, the evidence and arguments presented in this study should be taken as an invitation for further explorations into the connections between the acquisition of restricted collocations and exposure, as well as other related factors. Development of the NST corpus, the cloze test instrument and the plausibility metric methodology developed for this thesis will provide avenues for this to take place.

REFERENCES

Abraham, R. G., & Chapelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *Modern Language Journal*, 76, 468-479.

Adolphs, Svenja and Valerie Durow. (2004). Social-cultural intergration and the development of formulaic sequences. In N.Schmitt (ed.), *Formulaic Sequences* (pp. 127-51). Amsterdam: John Benjamins.

Adolphs, S. and Schmitt, N. 2003. "Lexical coverage of spoken discourse". *Applied Linguistics* 24, 4: 425-438.

Aisenstadt, E. 1981. Restricted collocations in English lexicology and lexicography. *ITL: Review of Applied Linguistics*, 53, 53-61.

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-227.

Alderson, J. C. (1980). Native and non-native speaker performance on cloze test. *Language Learning*, 30, 59-76.

Alis Puteh (2010). The language medium policy in Malaysia: A plural society model? *Review of European Studies*, 2(2), 192-200.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications* (pp. 101-22). Oxford: Oxford University Press.

Alternberg, B. and Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics* 22, 2, 173-94.

Ang, Leng Hong, Hajar Abdul Rahim, Tan , Kim Hua and Khazriyati Salehuddin, (2011) Collocations in Malaysian English learners' writing: a corpus-based error analysis. *3L; Language,Linguistics and Literature,The Southeast Asian Journal of English Language Studies*, 17 (special issue). pp. 31-44.

Arshad Abdul Samad, Fauziah Hassan, Jayakaran Mukundan, Ghazali Kamarudin, Sharifah Zainab Syd. Abd. Rahman, Juridah Md. Rashid and Malachi Edwin Vethamani. (2002). *The English of Malaysian School Students (EMAS) corpus*. Serdang: Universiti Putra Malaysia press.

Asmah Hj. Omar. (1992). *The Linguistic Scenery in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Bahns, J. (1993). Lexical Collocations: A Contrastive View. *ELTJ, 47*(1), 56-63.

Bahns, J., & Eldaw.M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101-114.

Bao, Z. (2003). Social stigma and grammatical autonomy in non-native varieties of English. *Language in Society 32*, 23-46.

Bao, Z. (2006). Variation in Non-native Varieties of English. In E.-. Keith Brown & Chief) (Eds.), *Encyclopedia of Language & Linguistics* (2nd ed., Vol. 13, pp. 377-380). Oxford: Elsevier.

Baskaran, L. (1987). *Aspects of Malaysian English Syntax*. Unpublished PhD Thesis. University of London.

Baskaran, L. (1994). The Malaysian English Mosaic. *English Today, 37*(10), 27-32.

Baskaran, L. (2002). *Whose English? Whose? What Standard? The Role of English in Malaysia*. Paper presented at the The International Conference IPBA.

Bayatee, D. (2007). *Cohesion and Coherence in English Essays Written by Malaysian and Thai Medical Students*. Paper presented at the Southern Thailand English Language Teaching/Cultural Change Conference.

Becker, J. D. (1975). *The Phrasal Lexicon*. Boston: Bolt Beranek and Newman, Inc.

Bennet, G.R. (2010). Using Corpora in the language learning classroom:Corpus linguistics for teachers. Michigan. *ELT*. Retrived online 10 September 2013 at http://www.press.umich.edu/titleDetailDesc.do?id=371534

Benson, M. (1985). Collocations and Idioms. In R. Ilson (Ed.), *Dictionaries, Lexicography and Language Learning* (pp. 61-68). Oxford: Pergamon Press.

Benson, M. (1990). Collocations and general-purpose dictionaries. *International Journal of Lexicography, 3*(1), 23-35.

Benson, M., Benson, E. & Ilson, R. (1986). *The BBI Combinatory Dictionary of English.* Amsterdam: John Benjamins.

Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3-43.

Biskup, D. (1992). L1 influence on learners' rendering of English collocations: A Polish/German empirical study. In A. a. Bejoint (Ed.), *Vocabulary and Applied LInguistics*. London: MacMillan.

Boers, F., Lindstromberg, S., and Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 acquisition. *Vigo International Journal of Applied Linguistics,* 11, 41-62.

Boers, F., Demecheleer, M., Coxhead, A., and Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research.* 18 (1), 54-74.

Boers, F., Demecheleer, M., and Eyckmans, J. (2004). Etymology elaboration as a strategy for learning figurative idioms. In P. Bogaards and B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 53-78). Amsterdam, the Netherlands: John Benjamins.

Boers, F., J. Eyckmans, J. Kappel, H. Stengers & M. Demecheleer (2006) Formulaic Sequences and perceived oral proficiency: Putting a lexical approach to the Test. *Language Teaching Research*, 10, 245-261.

Bolinger, D., & Sears, D.A. . (1968). *Aspects of language*. New York: Harcourt Brace Jovanovich Inc.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum, 1*, 1-14.

Botley, S., & Dillah, D. (2007). Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, 3, 74-93.

Boudah, D.J. (2011). *Conducting Educational Research*. Sage Publication.

Brown, J.D. (1989). Cloze item difficulty. *JALT Journal*, 11, 46-67.

Brown, J.D. (1993). What are the characteristics of natural cloze tests? *Language Testing*, 10, 93-116.

Carter, R. (1987). *Vocabulary: Applied linguistics perspective*. London: Allen and Unwin Ltd.

Channell, J. (1981). Applying Semantic theory to vocabulary teaching. *ELT Journal, 35*(115-122).

Church, K. W. and Hanks, P. (1991). Word association norms, mutual information and lexicography. *Computational Linguistics*, 16, 22-29.

Cléments, R. (1980). Ethnicity, contact and communicative competence in second language. In H. Giles, W. P. Robinson & P. M. Smith (eds.*), Language: social psychological perspectives*, 147-54.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*. 11.3, 38-63.

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.) Hillsdale, NJ: Erlbaum.

Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics, 3*, 239-266.

Coulmas, F. (1981). *Conversational Routine*. The Hague: Mouton.

Cowie, A. P. (1991). Multiword units in Newspaper Language. In S.Granger (Ed.), *Perspective on the English Lexicon: A Tribute to Jacques van Roey* (pp. 101-116).

Cowie, A. P. (1992). Multiwords lexical units and communicative language teaching. In P. J. L. Arnaud and H. Béjoint (eds): *Vocabulary and Applied Linguistics*. Houndmills: Macmillan, pp. 1-12

Cowie, A. P. (1994). Pharesology. In R. E. A. a. J.Simpson (Ed.), *The Encyclopedia of Language and Linguistics* (pp. 3168-3171). Oxford: Pergamon Press.

Cowie, A. P. (1998). Introduction-Phraseology-Theory, Analysis and Applications. In A.P.Cowie (Ed.), *Phraseology-Theory, Analysis and Applications* (pp. 1-20). Oxford: Clarendon Press Oxford.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34.2, 213-238

Coxhead, A. (2006). *Essentials of teaching academic vocabulary*. Boston, MA: Heinle Cengage.

Coxhead, A. (2011). A new academic word list 10 years on: research and teaching implications. *TESOL Quarterly* 45.2, 355-362

Crystal, D. (1987). How many words? *English Today*, 12.

Crystal, D. (Ed.) (1997). *A Dictionary of Linguistics and Phonetics.* Cambridge,MA: Blackwell Publishers.

Cutting, J. and Bock, K. (1997). That's the way the cookie bounces: syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition* 25: 1, 57–71.

Dagut, M. and Laufer, B. 1985. Avoidance of phrasal verbs-a case for contrastive analysis. *Studies in Second Language Acquisition*, 7: 73-80.

De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair and M. Hundt, M. (eds), *Corpus Linguistics and Linguistics Theory* (pp. 51-68). Amsterdam: Rodopi.

De Cock, S. (2004) Preferred sequences of words in NS and NNS speech. *Bell 2*: 225-246.

Dell, G.S. (1986) A Spreading –Activation Theory of Retrieval in Sentence Production. *Psychological Review.* Vol. 93, No. 3, 283-321.

Dörnyei, Z., & Katona, L. (1993). The C-test: A teacher friendly way to test language proficiency. *English Teaching Forum*, 31/1, 34-36.

Dörnyei, Z. And Skehan, P. (2003). Individual differences in second language learning. In *The Handbook of Second Language Acquisition*, C.J. Doughty and M.H. Long (eds), 589-630. Oxford: Blackwell

Dörnyei, Z.,Durow, V. and Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (ed.*), Formulaic Sequences* (pp. 87-106). Amsterdam: John Benjamins.

Durrant, P. (2008). *High frequency collocations and second language learning.* (Unpublished doctoral dissertation). The University of Nottingham, Nottingham.

Durrant, P., and Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics, 47*(2), 157-177.

Durrant, P., and Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 28, 163-188.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: OUP

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20,29-62.

Escaip, V. (2008). *The acquisition of phrasal vocabulary by non-native speakers of Spanish.* Unpublished Master's thesis. University of Canterbury.

Eyckmans, J., Boers, F., and Stengers, H. (2007). Identifying chunks: Who can see the wood for the trees? *Language Forum*, 33, 85-100.

Fellbaum, C. (2007). *Idioms and collocations: corpus-based linguistic and lexicographic studies*. London: Continuum.

Fernando, C. (1996). *Idiom and Idiomaticity*. Oxford: Oxford University Press.

Firth, J. R. (1957). 'A Synopsis of Linguistic Theory, 1930-1955' In G. Kennedy (Ed.), *An Introduction to Corpus Linguistics*. New York: Addison Wesley Longman,Inc.

Fishman, J. A. (1968). 'Nationality-Nationalism and Nation-Nationism'. In C. A. F. a. J. D. G. Joshua A.Fishman (Ed.), *Language Problems of Developing Nations*: John Wiley.

Fitzpatrick, Tess and Christina Izura (2011). Word association in L1 and L2. *Studies In Second Language Acquisition* 33: 373-398.

Fitzpatrick, Tess. (2012). Tracking the changes. *The Language Learning Journal*, 40(1); 81-98.

Foster, P. (2001). Rules and routines: A consideration of their role in task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75-93). Harlow: Longman.

Garside, R**.** (1987). The CLAWS Word-tagging System. In: R. Garside, G. Leech and G. Sampson (eds), *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.

Gaudart, H. (1997, 18-19 August,1997). *Malaysian English, Can or Not?* Paper presented at the English is An Asian Language: The Malaysian Context, Kuala Lumpur.

Geyken, A., Sokirko, A., Rehbein, I., and Fellbaum, C. (2004). *What is the optimal corpus size for the study of idioms?* Annual meeting of the German Linguistic Society, Mainz

Glaser, R. (1986). Phraseology der Englischer Sprache. In T. Miyakoshi (Ed.), *Investigating ESL learners' lexical collocations; The Acquisition of Verb + Noun Collocations by Japanese Learners of English.A doctoral dissertation*. USA.

Glaser, R. (1998). The stylistic potential of phraseolical units in the light of genre analysis. In A. P. Cowie (Ed.), *Phraseology-Theory, Analysis and Applications*. United States: Clarendon Press Oxford.

Goulden, R., Nation, I.S.P. and Read, J. (1990) How large can a receptive vocabulary be? *Applied Linguistics* 11, 4: 341-363.

Granger, S. (1998). Prefabricated Patterns in Advanced EFL Writing; Collocations and Formulae. In A. P. Cowie (Ed.), *Phraseology- Theory, Analysis and Applications*. Oxford: Clarendon Press.

Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.

Gupta, A. F. (1998). *Singapore Colloquial English? Or deviant standard English?* . Paper presented at the SICOL-Second International Conference on Oceanic Linguistics, Canberra.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. Lund: Lund University.

Hafriza Burhanudeen (2006). *Language and Social Behaviour: Voices from the Malay World*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In J. C. C. C.E. Bazell, M.A.K. Halliday & R.H.Robinson (Ed.), *In memory of J.R.Firth* (pp. 148-162). London: Longman.

Halliday, M. A. K. and Ruqayia Hassan 1976. *Cohesion in English*. London,

Longman.

Hancioğlu, N., S. Neufeld & J. Eldridge (2008). Through the looking glass and into the land of lexico-grammar. *English for Specific Purposes* 27, 459-479.

Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language* 20, 27-41.

Ho Mian-Lian and John, T. Platt (1993) Dynamics of a Contact Continuum: Singaporean English. Oxford Studies in *Language Contact.* Oxford: Clarendon Press.

Horst, M., Cobb, T., and Maera, P. (1998). Beyond *A Clockwise Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207-223.

Howarth, P. A. (1996). *Phraseology in English Academic Writing:Some Implications for Language Learning and Dictionary Making*. Tubingen: Max Niemeyer Verlag.

Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics, 19*(1), 24-44.

Jackendoff, R. (1995). Phrasemes in language and phraseology in linguistics. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum.

Jackendoff, R. (1997). 'Twisting' the night away, *Language*, 73:534-59.

Jackson, H. (1988). *Words and their meaning*. London: Longman.

Jafapur, A. (1995). Is C-testing superior to cloze? *Language Testing*, 12 (2), 194-216.

Kachru, B. B. (1985). Standards, Codifications and Sociolinguistics Realism:The English Language in the Outer Circle. In R. Q. H. Widdowson (Ed.), *English in the World,Teaching and Learning Language and Literature* (pp. 11-30). Cambridge & New York: Cambridge University Press.

Kachru, B. B. (1992). *The other tongue-English across culture*. Oxford: Pergamon Press.

Kachru, B. B. (1992). The Second Diaspora of English'. In T. W. M. C. T. Scott (Ed.), *English in its Social contexts: Essays in historical Sociolinguistics* (pp. 230-252). New York: Oxford University Press.

Kachru, B. B. (1996). *Opening Borders with World Englishes: Theory in the classroom.* Paper presented at *JALT, Hiroshima.*

Kamariah Yunus & Su'ad Awab. (2011) . Collocational competence among Malaysian Law undergraduate students. *Malaysian Journal of ELT Research*, 7(1), 151-202.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: Longman.

Kilgarriff, A. (1995) *BNC Database and Word Frequency Lists* [cited 1 August 2012 Available from www.kilgarriff.co.uk/BNC_lists/lemma.al].

Kirkpatrick, A. (2007) Implications for English Language Teaching. In A. Kirkpatrick *World Englishes: Implications for International Communication and English Language Teaching*. Pp 184-197. Cambridge University Press: Cambridge

Kjellmer, G. (1982). Some problems relating to the study of collocations in the Brown Corpus. In S.Johansson (Ed.), *Computer Corpora in English Language Research* (pp. 25-33). Bergen: Norwegian Computing Centre for the Humanities.

Knowles, G and Zuraidah Mohd Don. (2004). Introducing MACLE: The Malaysian Corpus of Learner English. *1st National Symposium of Corpus Linguistics and Foreign Language Education*, South China Normla University, Guangzhou, China, 10-14 October 2004.

Kobayashi, M. (2002). Cloze test revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86 (4), 571-586.

Kormos, J. (2006). *Speech production and second language acquisition. Cognitive sciences and second language acquisition.* Lawrence Erlbaum Associates, Mahwah:NJ.

Koya, T. (2004). Collocation Research Based on Corpora Collected from Secondary School Rextbooks in Japan and in the UK. *Dialogue,TALK, 3*(7-18).

Koya, T. (2005). *The acquisition of basic collocations by Japanese learners of English.* Unpublished doctoral dissertation. Waseda University.

Koya, T. (2006). 'What is the reality of collocation use by native speaker of English? *Dialogue,TALK, 5*, 1-18.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

Kuiper, K. And Haggo, D. C. (1984). Livestocks auctions, oral poetry and ordinary language. *Language in Society*, 13, 205-34.

Kuiper, K., & Austin., J.P.M. (1990). 'They're off and racing now: the speech of the New Zealand race caller' In A. B. J.Holmes (Ed.), *New Zealand Ways of Speaking English* (pp. 195-220). Bristol: Multilingual Matters.

Kuiper, K. (1996). *Smooth Talkers*. Mahwah, NJ: Erlbaum.

Kuiper, K., McCann, H., Aitchison, T. and van der Veer, K. (2003). SAID: *A Syntactically Annotated Idiom Database*. Linguistics Data Consortium, University of Pennsylvania.

Kuiper, K. &. A. Kuiper. (2003). Constructing Vocational Aspirations Linguistically. *Globalisation, Societies and Education, 1*(2), 223-240.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences: acquisition, processing and use* (pp. 37-54). Amsterdam: John Benjamins Publishing Company.

Kuiper, K., van Egmond, M-E., Kempen, G. M. and Sprenger, S. (2007). Slipping on superlemmas: multiword lexical items in speech production. *The Mental Lexicon*, 2: 3, 313-57.

Kuiper, K. (2009). *Formulaic Genres*. London: Palgrave-Macmillan.

Kuiper, K., Columbus,G.,& Schmitt,N. (2009). Acquiring phrasal vocabulary In S. Foster-Cohen (Ed.), *Advances in language acquisition*. Basingstoke: Palgrave Macmillan.

Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75, 440-448.

Laufer, B. (2000). Avoidance of idioms in a second language: the effect of L1-L2 degree of similarity. *Studia Linguistics*, 54, 186-96.

Laufer, B. (2001). Reading, word-focused activities and incidental vocabulary acquisition in a second language. *Prospect*, 16, 44-54.

Laufer, B., and Waldan, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61, 647-672.

Leech, G., Garside, R., and Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)* Kyoto, Japan, pp 622-628.

Levelt, W. J. M. (1989) *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. and Meyer, A.S. (2000). Word for word: multiple lexical access in speech production. *European Journal of Cognitive Psychology*, 12: 4, 433-52.

Lewis, M. (1997). *Implementing the lexical approach.* Hove, UK: LTP

Li, J., and Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.

Li, J., and Schmitt, N. (2010). The development of collocations use in academic texts by advanced L2 learners: a multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 23-46). London/ New York: Continuum.

Lindstromberg, S., and Boers, F. (2008a). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, 29, 200-222.

Lindstromberg, S., and Boers, F. (2008b). Phonemic repetition and the learning of lexical chunks: The mnemonic power of assonance. *System*, 36, 423-436.

Lombard, R. J. (1997). *'Nonnative speaker collocations: A Corpus-Driven Characterization from the writing of native speakers of Mandarin.* University of Texas at Arlington.

Makarova, E. (2010).Teaching Culture in English Classroom. Retrieved January 11, 2013 from http://www.mymla.org/files/icmm2010_papers/ICMM2010_p22.pdf.

Marlyna Maros, and Tan , Kim Hua and Khazriyati Salehuddin. (2007). Interference in learning English: Grammatical errors in English essay writing among rural Malay secondary school students in Malaysia. *Journal e-BANGI: Jurnal Sains Sosial dan Kemanusiaan*, 2 (2). 1- 15.

Masoomeh, Y., Yap Ngee Thai, Mardziah Hayati Abdullah, & Tan Bee Hoon (2012). The influence of cross-linguistic similarities on L2 idiom production. *3L: The Southeast Asian Journal of English Language Studies*, 18 (4): 91-107.

McEnery, T. and Wilson, A. (1996). *Corpus Linguitics*. Edinburgh University Press

Mel'cuk, I. (1995). Phrasemes in language and phraseology in linguistics. In M. Everaert, E. van der Linden, A. Schenk & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives*. Hillsdale, NJ: Erlbaum.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32, 129-148.

Miyakoshi, T. (2009). *Investigating ESL learners' lexical collocations; The Acquisition of Verb + Noun Collocations by Japanese Learners of English.* Unpublished doctoral dissertation. Hawai'i University, USA.

Mitchell, T. F. (1971). Linguistic 'goings on': Collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum*, 2 (1), 35-69.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus Based Approach*. Oxford: Clarendon Press.

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt and M. McCarty (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 40-63), Cambridge: Cambridge University Press.

Morais, E. (1997, 18-19 August 1997). *Talking in English but Thinking like a Malaysian: Insights from a Car Assembly Plant.* . Paper presented at the English is an Asian Language :The Malaysian Context, Kuala Lumpur.

Nagy, W., Anderson, R.C, Schomer, M., Scott, J.A., and Stallman, A.C. 1989. Morphological families in the internal lexicon. *Reading Research Quarterly* 24: 262-282.

Nair-Venugopal, S. (2001). The Sociolinguistics of choice in Malaysian Business Settings. *International Journal of Soc. Lang., 152*, 21-52.

Nation I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nation, P. and Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt and M. McCarthy (eds), *Vocabulary: Description, acquisition, pedagogy*. New York: Cambridge University Press.

Nation, I.S.P. (2000) Most Frequent Words [cited August 2012. Available on www.edict.biz/textanalyzer/wordlists.htm].

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language* Review 63/1, 59-82.

Nation I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching* 44.4, 529-539.

Nattinger, J. R and DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nesselhauf, N. (2003). Transfer at the locutional level: An Investigation of German-speaking and French –speaking learners of English. In Tschichold (Ed.), *English core linguistics Essays in honour of D.J. Allerton* (pp. 269-286). Bern: Peter Lang.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Newsome, G. (2005). *Estimating the size of the phrasal E-lexicon of aircraft maintenance engineers* (Unpublished master's thesis). University of Canterbury.

Ooi, V. B. Y. (1998). *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.

Ooi, V. B. Y. (2000). Asian or western Realities-Collocations in Singaporean-Malaysian English. *Corpus Galore* (pp. 73-89 (or 194-212)): Rodopi.

Palmer, H. E. (1933). Second Interim Report on English Collocation. In G. Kennedy (Ed.), *An Introduction to Corpus Linguistics.New York:Addison Wesley Longman,Inc.* Tokyo: Institute for Research in English Teaching.

Pavlenko, A. (1996). Bilingualism and cognition: concept in the mental lexicon. In Pavlenko, A., & Salaberry, R. (Eds.), *Papers in Second Language Acquisition and Bilingualism* (pp. 49-85). Ithaca: Cornell University Press.

Pawley, A. and Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (eds), *Language and Communication* (pp. 191-226). London: Longman.

Pawley, A. (1985). On speech formulas and linguistic competence. *Lenguas Modernas*, 12: 84-104.

Pawley, A. S. (1991). How to talk cricket. In Blust, R. *Currents in Pacific Linguistics: Papers in Austronesian Languages and Ethnolinguistics in Honour of George W. Grace*, pp. 339-368). Honolulu: Pacific Linguistics.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.

Pillai, S. and K., Fauziah. (2006). *The Variety of Malaysian English Used in Radio Advertisements.* Paper presented at The 8th English in Southeast Asia Conference, Malaysia.

Pillai, S. (2006). Malaysian English As a First Language. In M. K. David (Ed.), *Language Choices and Discourse of Malaysian Families (pp..61-75).* Petaling Jaya: SIRD.

Platt, J. T. (1975). The Singapore English speech continuum and its basilect 'Singlish' as a 'creoloid'. *Anthropological Linguistics*(17), 363-374.

Platt, J. T. and Weber, H. (1980). *English in Singapore and Malaysia:Status, features, functions*. Oxford: Oxford University Press.

Platt, J. T., Weber, H. & Ho, M. L. (1983). *Singapore and Malaysia: Varieties of English Around the World* (Vol. 4). Amsterdam: John Benjamin Publishing Company.

Plunkett, K. (1993). Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language*, 20, 43-60

Rafidah Kamarudin. (2013*). A Study on the Use of Phrasal Verbs by Malaysian Learners of English.* Unpublished doctoral dissertation. University of Birmingham.

Randall, M. 1980. Word association behavior in Learners of English as a foreign language. *Polyglot 2*, fiche2.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal 19*, 12-25.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition,* 21, 589-619.

Saadiyah Darus & Kaladevi Subramaniam. (2009). Error analysis of the written English essays of secondary school students in Malaysia: A case study. *European Journal of Social Sciences, 8* (3), 483-495.

Saragi, T., Nation, I. S. P., and Meister, G.F. (1978). Vocabulary learning and reading. *System*, 6, 72-78.

Sawyer, M., and Ranta, L. 2001. Aptitude, individual differences, and instructional design. In *Cognition and Second Language Instruction*, P. Robinson 9ed.), 319-353. Cambridge: CUP.

Scarcella, R.C. (1979). "Watch up!" : A study of verbal routines in adult second language performance. *Working Papers on Bilingualism*, Vol 19, 79-90.

Scheiner, E. W. (2003). The dynamics of New Englishes:From a dynamic constructio to a dialect birth. *Language 79*(2), 233-281.

Schmitt, N. 1998. Quantifying word association responses: what is native-like? *System* 26: 389-401.

Schmitt, N., (2000) *Vocabulary in Language Teaching*. Cambridge University Press.United Kingdom.

Schmitt, N., Schmitt, D., and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*(1), 55-88.

Schmitt, N. And Carter, R.2004. 'Formulaic sequences in action: An Introduction' in Schmitt, N. (ed.): *Formulaic sequences*. Amsterdam: John Benjamin.pp.1-22

Schmitt, N. Dörnyei, Z., Adolphs, S. and Durow, V. (2004) Knowledge and acquisition of formulaic sequences. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 127-51). Amsterdam: John Benjamins.

Schmitt, N., Grandage, S. And Adolph, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N.Schmitt (ed.), *Formulaic Sequences* (pp. 127-51). Amsterdam: John Benjamins.

Schmitt, N. (2005). Formulaic language: fixed and varied. *ELIA* 6, 13-39

Schmitt, N. (2010). *Researching Vocabulary-A Vocabulary Research Manual*. London: Palgrave Macmillan.

Schmitt and Schmitt, 2012 A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 1-20

Shameem Rafik-Galea, S & Mohd Salleh Hj. Hassan. (2003). Bilingualism in Malaysia Corporate Communication. *Studies in Foreign Language Education*.. 9pp. 91-112.

Shin, D. (2007). The high frequency collocations of spoken and written English. *English Teaching*, 62/1: 199-218

Shin, D. and Nation, P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62/4: 339-348

Sinclair, J. (1966). Beginning the study of lexis. In J. C. C. C.E. Basell, M.A.K. Halliday, & R.H.Robins (Ed.), *In Memory of J.R.Firth* (pp. 410-430). London: Longman.

Sinclair, J. (1991). *Corpus,Concordance and Collocation*. Oxford: OUP.

Siyanova, A. and Schmitt, N. (2007). Native and non-native use of multiword Versus one-word verbs. *International Review of Applied Linguistics in Language Teaching* 45 (2): 119-139.

Siyanova, A. and Schmitt, N. (2008). L2 learners production and processing of collocation: a multi-study perspective. *The Canadian Modern Language Review*, 64 (3), 429-458.

Skehan, Peter (1998*). A Cognitive Approach to Learning Language.* Oxford: Oxford University Press.

Smadja, F. (1993). 'Retrieving Collocations from Text:Xtract' *Association for Computational Linguistics, 19*(1), 143-177.

Sökmen, A. (1993) Word association results: a window to the lexicons of ESL students. *JALT Journal* 15/2, 135-50

Spöttl, C. and McCarthy, M (2004). Comparing knowledge of formulaic sequences across L1, L2, L3 and L4. In N. Schmitt (ed.), *Formulaic sequences.* Amsterdam: John Benjamins. Pp. 190-225

Sprenger, S., Levelt, W. J. M. and Kempen, G. (2006). Lexical access during theproduction of idiomatic phrases. *Journal of Memory and Language*, 54, 161–84.

Stengers, H., F. Boers, J. Eyckmans and A. Housen (2010) Does chunking foster chunk uptake? In S. De Knop, F. Boers and A. De Rycker (eds.), *Fostering Language Teaching Efficiency through Cognitive Linguistics*, 239-256 Berlin/ New York: Mouton de Gruyter.

Stubbs, J.B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *The Modern Language Journal*, 58 (5/6), 239-241.

Su'ad Awab (1994). A Corpus of Proposals: A Description and Implications. *Jurnal Bahasa Moden, 8,* 79-89.

Tremblay. A., Bayern, R.. H., Derwing, B., and Libben, G. (2008). *Frequency and the processing of multiword strings: A behavioural and ERP study.* Paper presented at the conference of the Canadian Linguistics Association , Université de la Colombie- Britannique, May 31 2008. *http://ocs.sfu.ca/fedcan/index.php/cla/acl-cla2008/paper/viewFile/219/152*

Vine, E.W. (2011). High frequency multifunctional words: accuracy of word-class tagging. *Te Reo*, 54, 71-82.

Waring, R., and Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading graded readers? *Reading in a Foreign Language*, 15, 1-27.

Webb, S., Newton, J., and Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63, 91-120.

Webb, S., and Kagimoto, E. (2011). Learning collocation: Do the number of collocates, position of the node word, and synonym affect learning? *Applied Linguistics*, 32, 259-276.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30, 79-95.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Reading in Foreign Language,* 19, 120-136.

Willis, J. D. (1990). *The lexical syllabus*. London: Collins.

Wray, A. (1999). Formulaic language in leaners and native speakers. *Language Teaching*, 32 (4): 213-31

Wray, A. (2002), *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. and Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, 20, 1-28.

Yamashita, J., and Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44, 647-668.

Corpora
British National Corpus
NST Corpus

Dictionary
The idiom list in *Syntactically Annotated Dictionary of Idioms (SAID)*, a dictionary of idioms (including phrasal verbs) (Kuiper et al., 2003).

Verb frequency list
Kilgarriff's lemmatized BNC frequency list.

Software
Wordsmith Tools 5.0. Scott, M. (2010). Oxford University Press.

Tagger
CLAWS POS tagger. UCREL. University of Lancaster.

Database / Archive
Phrases in English (PIE). Derived from British National Corpus (BNC).
News Straits Times Archive. News Straits Times Press, Malaysia.

Appendix A

Hasliza Abd Halim

From: azlee@nstp.com.my
Sent: Tuesday, 24 May 2011 10:25 p.m.
To: Hasliza Abd Halim
Cc: yazid@nstp.com.my
Subject: Re: FW: PhD research-Request for a year's copy of news and current affairs from NST (2010)

Dear Hasliza,

Regarding your request to Datuk Syed Nadzri on the above subject, you can view our website at http://www.nib.com.my
Go to **Pools** - select/ click **Text Archive**
- click **Publication:** select **New Straits Times**
- go to **From:** -- key-in or select Jan 1, 2010
- go to To: -- key-in or select Dec 31, 2010
- **Search**
- click at **Headline** to view the full article/s

You can copy and paste the article/s without any charges (7 years back only).

Regards
Mohd Azlee Bin Abd Aziz
Resource Centre
The New Straits Times Press (Malaysia) Berhad
Balai Berita
31 Jalan Riong
59100 Kuala Lumpur
Malaysia
Phone: +603 2056 9375

-----Syed Nadzri b Syed Harun/NST/NSTP wrote: -----

To: azlee@nstp.com.my
From: Syed Nadzri b Syed Harun/NST/NSTP
Date: 05/24/2011 05:32PM
Cc: yazid@nstp.com.my
Subject: FW: PhD research-Request for a year's copy of news and current affairs from NST (2010)

-----Forwarded by Syed Nadzri b Syed Harun/NST/NSTP on 05/24/2011 05:27PM -----

To: syedn@nstp.com.my
From: Hasliza Abd Halim <hasliza.abdhalim@pg.canterbury.ac.nz>
Date: 05/12/2011 09:57AM
Subject: FW: PhD research-Request for a year's copy of news and current affairs from NST (2010)

(See attached file: Letter to the editor(NST)1.doc)

From: Hasliza Abd Halim
Sent: Thu 5/12/2011 1:35 PM
To: news@nstp.com.my
Cc: Hasliza Abd Halim

1

Subject: PhD research-Request for a year's copy of news and current affairs from NST (2010)

Dear Dato'/Datuk/Datin/Sir/Madam,

I'm Hasliza Abdul Halim a PhD at the University of Canterbury. I'm conducting a research on student's acquisition of phrasal vocabulary. To conduct this research I need to have a year's copy of news and current affairs in text only-form. We are giving the honour by naming the corpus as 'NST corpus' . This copy is very vital to this research project as I need a large corpus to work with otherwise the frequencies are not reliable.

Together with this email, I attach 2 letters written by my supervisor, Prof Emeritus Kon Kuiper. The 1st letter was mailed to you in February, and the second one is the follow-up letter.

I would like to thank you for your kind consideration.

regards,
Hasliza

[attachment "Letter to the editor(NST)1.doc" removed by Mohd Azlee b Abd Aziz/EMEDIA/NSTP]

The NSTP Group is part of the Media Prima portfolio of companies and is the nation's longest running and largest newspaper publishing company, reaching five million Malaysians daily. While our heritage keeps us rooted, our current transformation programme is designed to take partnership with all our stakeholders to new heights.

This email and any attachments herein are confidential and intended solely for the use of the addressee. The NSTP Group of Companies gives no warranty as to the entirety, accuracy, or security of this email nor guarantees that it is virus free.

2

College of Arts
School of Languages and Cultures
Tel: +64 3 364 2184 or +64 3 364 2556, Fax: +64 3 364 2598, Email: lanc@canterbury.ac.nz

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

11th May 2011

Y.Bhg. Dato' Syed Nadzri Syed Harun
Group Editor, NSTSB / Pengarang Kumpulan
NSTSB
Balai Berita,
31 Jalan Riong,
59100 Kuala Lumpur,
Malaysia.

Dear Y.Bhg. Dato' Syed Nadzri Syed Harun
I am writing to follow up a letter I wrote to you in February requesting your assistance with the preparation of a corpus of formal written Malaysian English in the form of a year's copy of news and current affairs from the *New Straits Times*.
I wonder if you could let me know whether you and your staff have been able to consider my request and what your decision is.

Yours sincerely

Koenraad Kuiper, MA. PhD, DSc, Professor Emeritus

14th February 2011

Y.Bhg. Dato' Syed Nadzri Syed Harun
Group Editor, NSTSB / Pengarang Kumpulan
NSTSB
Balai Berita,
31 Jalan Riong,
59100 Kuala Lumpur,
Malaysia.

Dear Y.Bhg. Dato' Syed Nadzri Syed Harun
I am writing to you as the Group Editor of Malaysia's leading English language newspaper soliciting your assistance. Around the world there is considerable interest in "New Englishes" of which Malaysian English is one. It is little studied. Many varieties of English are now studied using large corpora, machine readable collections of texts which can then be annotated and searched for various properties of the language in which the texts are written. One such corpus is currently being created in Malaysia. It is the Malaysian section of the International Corpus of English. It is a so-called balanced corpus containing standard word counts for various kinds of language use such as formal written varieties, spoken varieties of various kinds. It is, however, relatively small containing only a million words when it is completed.
Many other corpora have a sizable component of newspaper copy since that is machine readable and easy to convert into text-only form to allow it to be searched. Since the New Straits Times is written in formal English it provides a very good source of information on the variety of English used for formal writing in Malaysia. It is important to understand this variety since it is what the education of young Malaysians is aimed at inculcating.
I have a doctoral student who is working on a project to see how successful the English language education of Malaysian students is at having them acquire the idioms which are so important in having someone sound and write like a fluent native speaker. However, at the moment, the only norms we have to go on are other varieties of English than that which is becoming the standard in Malaysia. We are particularly interested in the frequency with which words occur rather than the vocabulary of Malaysian English itself. This can now be ascertained with various kinds of software but requires a large corpus to work with otherwise the frequencies are not reliable.
All this by way of preamble to a request for a year's worth of news and current affairs copy, to be named the "New Straits Times corpus" with full acknowledgement, so we can find out whether words are used with the same kinds of frequencies as they are in, for example, standard British English. To be usable this should be in text only-form. The PhD project would then be able to use reliable frequency data to see how well Malaysian students are learning Malaysian English. The corpus could also be used by Malaysian researchers interested in Malaysian English. It would be our intention to

make it available for that purpose after it has been annotated with part of speech labels.

Yours sincerely

Professor Koenraad Kuiper, MA. PhD, DSc, Professor Emeritus

Appendix B

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON

FACULTY OF HUMANITIES AND SOCIAL SCIENCES
TE WAHANGA ARONUI

FACULTY RESEARCH COMMITTEE
RESEARCH FELLOW AND POSTGRADUATE STUDENT POLICY

APRIL 2011

Dear Hasliza,

Congratulations on your successful application for a Chair of Malay Studies Research Grant. The total amount awarded is $4265 and should be used in 2011. The following expenditure has been approved;

Flights	$2900
15 x per diem @ $91	$1365

Please note that the per diem is a contribution towards accommodation, food, airport transfers, internal transport - see attached policy Section 9. You are reminded that you must keep all receipts relating to your per diem to show how it has been spent. All the receipts must be attached to your reimbursement claim form when you return to NZ. The receipts and attached claim form must be sent to Joan Harvey, Finance Advisor, MY 315, Victoria University, P O Box 600, Wellington.

As you are a Canterbury University student, the grant (being internal) cannot be set up in your name.
Prof Wan Rafaei has kindly agreed to it being set up in his name. This means that all your travel must be booked through the Victoria University travel provider - Atlantic Pacific American Express (APX) www.apx.co.nz
Please contact Paul Noir 04 801 2646 once you receive confirmation from me that your grant has been set up. APX have direct access to grant funds and will charge the flight cost to your grant.

You are welcome to contact me if you need further information.

Best wishes with your research.

Cheryl McGettigan
Research Committee Administrator
Faculty of Humanities and Social Sciences
Victoria University
WELLINGTON

Phone 463 5532

Dear Hasliza,

Congratulations on your successful application for a Chair of Malay Studies Research Grant. The total amount awarded is $4265 and should be used in 2011. The following expenditure has been approved;

Flights	$2900
15 x per diem @ $91	$1365

Please note that the per diem is a contribution towards accommodation, food, airport transfers, internal transport - see attached policy Section 9. You are reminded that you must keep all receipts relating to your per diem to show how it has been spent. All the receipts must be attached to your reimbursement claim form when you return to NZ. The receipts and attached claim form must be sent to Joan Harvey, Finance Advisor, MY 315, Victoria University, P O Box 600, Wellington.

As you are a Canterbury University student, the grant (being internal) cannot be set up in your name.
Prof Wan Rafaei has kindly agreed to it being set up in his name.
This means that all your travel must be booked through the Victoria University travel provider - Atlantic Pacific American Express (APX) www.apx.co.nz
Please contact Paul Moir 04 801 2646 once you receive confirmation from me that your grant has been set up. APX have direct access to grant funds and will charge the flight cost to your grant.

You are welcome to contact me if you need further information.

Best wishes with your research.

Cheryl McGettigan
Research Committee Administrator
Faculty of Humanities and Social Sciences
Victoria University
WELLINGTON

Phone 463 5532

Email: cheryl.mcgettigan@vuw.ac.nz

Dear Hasliza,

The system generated notification below, confirms the details of your approved grant. These are required when you book travel through APX (the University's travel provider) or when you wish to make a claim against the grant.

Attached is the document which explains the use of your grant and a reimbursement form which you may require. Please ensure that any reimbursements are applied for with receipts in hard copy and are signed off by Joan Harvey – Finance Adviser FHSS, the only staff member with authorisation to use this budget centre. All claims must be completed prior to the project funding end date indicated below.

Best wishes in completing this research.

Cheryl

Cheryl McGettigan
Faculty Research Committee Administrator

Phone 463 5532

Cheryl.mcgettigan@vuw.ac.nz

Office hours - Tues/Wed/Thurs 10 am – 2.30 pm

From: Workflow Mailer PROD [mailto:NoReplyTo@staff.vuw.ac.nz]
Sent: Friday, 24 June 2011 9:35 a.m.
To: Cheryl McGettigan
Subject: FYI: Grant Created in GMS by Grants Administrator

To	**McGettigan, Cheryl**
Sent	**24-JUN-11 09:34:32**
ID	**2350300**

THIS IS A SYSTEM GENERATED NOTIFICATION

Hello Cheryl

Please be advised that a Grants application has been setup in the GMS System.

Details of the Grants application are as follows;

Submitter Name	Donna Jamieson
Applicant Name	Wan Rafaei
Application Type	Internal Research Grant
Application Sub-Type	Research
Application Description	Malay Chair 2011 - Hasliza Halim PhD
Applicant School	Languages and Cultures
Application ID	109814

The project/award details are as follows;
Project Number: 109814
Award Number: 2350
Project Name: Rafaei, W - Malay Chair 2011
Total Budget: $4,265.00
Project Manager: Sekhar Bandyopadhyay
Project (Funding) Start Date: 23-JUN-2011
Project (Funding) End Date: 31-DEC-2011
Cost Centre: 3330

The above project and award number should be used on all expenditure, and this project is now available for query in the Grants Web Forms.

To log into the Grants Web Forms System, please click here.

If you have any queries then please contact the Grants Administrator on GMS-Help@vuw.ac.nz or Ph extn. 6723.

Regards
Grants Administrator

Appendix C

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Lynda Griffioen
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2011/11

18 April 2011

Hasliza Binti Abd Halim
Department of Languages, Cultures & Linguistics
UNIVERSITY OF CANTERBURY

Dear Hasliza

The Human Ethics Committee advises that your research proposal "Restricted verb-phrase collocations in standard and learner Malaysian English" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 13 April 2011.

Please ensure the second CLOZE test is submitted to the University of Canterbury Human Ethics Committee for review.

Best wishes for your project.

Yours sincerely

pp LGriffioen

Dr Michael Grimshaw
Chair, Human Ethics Committee

UNIT PERANCANG EKONOMI
Economic Planning Unit
JABATAN PERDANA MENTERI
Prime Minister's Department
BLOK B5 & B6
PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN
62502 PUTRAJAYA
MALAYSIA

Telefon : 603-8872 3333

Ruj. Tuan:
Your Ref.:

Ruj. Kami:
Our Ref.: UPE: 40/200/19/2821

Tarikh:
Date: 30 Jun 2011

Hasliza Abd Halim
No.645 Tmn Tunku Sarina
Lorong Diana 6
Bandar Darul Aman
Jitra 06000, Kedah.
Email: haslieza@uum.edu.my

APPLICATION TO CONDUCT RESEARCH IN MALAYSIA

With reference to your application, I am pleased to inform you that your application to conduct research in Malaysia has been *approved* by the **Research Promotion and Co-Ordination Committee, Economic Planning Unit, Prime Minister's Department.** The details of the approval are as follows:

Researcher's name : **HASLIZA ABD HALIM**

Passport No. / I. C No: **780202-02-5784**

Nationality : **MALAYSIAN**

Title of Research : **"RESTRICTED VERB – PHRASE COLLOCATIONS IN STANDARD AND LEARNER MALAYSIAN ENGLISH"**

Period of Research Approved: **3 YEARS**

2. Please collect your Research Pass in person from the Economic Planning Unit, Prime Minister's Department, Parcel B, Level 1 Block B5, Federal Government Administrative Centre, 62502 Putrajaya and bring along two (2) passport size photographs. You are also required to comply with the rules and regulations stipulated from time to time by the agencies with which you have dealings in the conduct of your research.

3. I would like to draw your attention to the undertaking signed by you that you will submit without cost to the Economic Planning Unit the following documents:

a) A brief summary of your research findings on completion of your research and before you leave Malaysia; and

b) Three (3) copies of your final dissertation/publication.

4. Lastly, please submit a copy of your preliminary and final report directly to the State Government where you carried out your research. Thank you.

Yours sincerely,

(**MUNIRAH ABD. MANAN**)
For Director General,
Economic Planning Unit.
E-mail: munirah@epu.gov.my
Tel: 88882809
Fax: 88883961

ATTENTION

This letter is only to inform you the status of your application and **cannot be used as a research pass**.

Cc:

Ketua Setiausaha
Bahagian Perancangan Penyelidikan & Dasar Pendidikan
Kementerian Pelajaran Malaysia
Aras 1-4, Blok E8, Kompleks Kerajaan Parcel E
Pusat Pentadbiran Kerajaan Persekutuan
62604 Putrajaya

Ketua Setiausaha
Kementerian Pengajian Tinggi Malaysia
Bahagian Perancangan dan Penyelidikan
Aras 3, Blok E3, Kompleks E.
Pusat Pentadbiran Kerajaan Persekutuan
62505 Putrajaya.
(up: Hjh. Raihanah Bt. Hj. Khudri)

Appendix D

Malay medium private schools in Malaysia-2011

BIL	NEGERI	NAMA SEKOLAH
1	PERAK	SEKOLAH RENDAH TENBY IPOH
2	PERAK	SEKOLAH RENDAH WESLEY METHODIST IPOH
3	SELANGOR	SEKOLAH RENDAH SRI CEMPAKA
4	SELANGOR	SEKOLAH RENDAH SRI KDU
5	PAHANG	SEKOLAH RENDAH SERI OMEGA
6	PAHANG	SEKOLAH SRI UTAMA
7	PAHANG	SEKOLAH RENDAH SRI ARA
8	PAHANG	SEKOLAH RENDAH SRI CAHAYA
9	KELANTAN	SEKOLAH RENDAH WADI SOFIA
10	KEDAH	SEKOLAH RENDAH AN-NAJAAH
11	N.SEMBILAN	SEKOLAH RENDAH BERSEPADU KEMAYAN

12	P.PINANG	SEKOLAH RENDAH SRI PINANG
13	P.PINANG	SEKOLAH RENDAH SRI TENBY
14	TERENGGANU	SEKOLAH RENDAH SRI UTAMA, K.TERENGGANU
15	K.LUMPUR	SEKOLAH RENDAH SATHYA SAI
16	K.LUMPUR	SEKOLAH RENDAH SRI DASMESH
17	K.LUMPUR	SEKOLAH RENDAH SRI SEMPURNA
18	K.LUMPUR	SEKOLAH SRI BUNGA RAYA (RENDAH)
19	K.LUMPUR	SEKOLAH SRI CEMPAKA
20	K.LUMPUR	SEKOLAH RENDAH SRI GARDEN
21	K.LUMPUR	SEKOLAH SRI UTAMA KUALA LUMPUR (RENDAH)
22	LABUAN	SEKOLAH SRI LABUAN
23	SABAH	SEKOLAH RENDAH ADVENT DAMAI
24	SABAH	SEKOLAH RENDAH ADVENT GAUR

25	SABAH	SEKOLAH RENDAH ADVENT MERABAU
26	SABAH	SEKOLAH RENDAH ADVENT PODOS
27	SABAH	SEKOLAH RENDAH ADVENT TAMBULURAN KUDAT
28	SABAH	SEKOLAH RENDAH ADVENT TAMPARULI
29	SABAH	SEKOLAH RENDAH ADVENT TENGHILAN
30	SABAH	SEKOLAH RENDAH SDA ADVENT SUNGOI KOTA MARUDU
31	SABAH	SEKOLAH RENDAH SDA GOSHEN
32	SABAH	SEKOLAH RENDAH SDA KELAWAT
33	SABAH	SEKOLAH RENDAH SDA RENGALAU
34	SABAH	SEKOLAH RENDAH SDA TAGAROH
35	SABAH	SEKOLAH RENDAH SERI INSAN
36	SABAH	SEKOLAH RENDAH SWASTA DATUK SIMUN FUNG
37	SABAH	SEKOLAH RENDAH AL-SALAM

38	SABAH	SEKOLAH RENDAH VISI
39	SARAWAK	SEKOLAH RENDAH LODGE
40	SARAWAK	SEKOLAH RENDAH SRI MAWAR
41	SARAWAK	SEKOLAH RENDAH SUNNY HILL
42	SARAWAK	SEKOLAH RENDAH TUNKU PUTRA
43	SARAWAK	SEKOLAH RENDAH SEKIM LAMBER
44	SARAWAK	SEKOLAH RENDAH SRI MULIA
45	PERAK	SEKOLAH TINGGI METHODIST IPOH
46	PERAK	SEKOLAH MENENGAH TENBY
47	SELANGOR	SEKOLAH MENENGAH SRI CAHAYA
48	SELANGOR	SEKOLAH MENENGAH SRI CEMPAKA CHERAS
49	SELANGOR	SEKOLAH MENENGAH SRI KDU
50	SELANGOR	SEKOLAH MENENGAH SRI SEDAYA

51	SELANGOR	SEKOLAH MENENGAH SERI SURIA
52	SELANGOR	SEKOLAH MENENGAH SRI ACMAR
53	SELANGOR	SEKOLAH BEACONHOUSE SRI INAI (MENENGAH)
54	SELANGOR	SEKOLAH SRI KUALA LUMPUR (MENENGAH)
55	SELANGOR	SEKOLAH MENENGAH SRI LETHIA
56	SELANGOR	SEKOLAH SRI MURNI (MENENGAH)
57	SELANGOR	SEKOLAH TINGGI METHODIST KLANG (MENENGAH)
58	SELANGOR	SAPURA SMART SCHOOL (MENENGAH)
59	SELANGOR	SEKOLAH MENENGAH TAMAN ILMU DAN BUDI
60	SELANGOR	SEKOLAH MENENGAH SRI BESTARI
61	SELANGOR	SEKOLAH MENENGAH SRI NOBEL
62	SELANGOR	SEKOLAH SRI TENBY
63	PAHANG	MAKTAB ADABI KUANTAN

64	PAHANG	SEKOLAH MENENGAH SALEHA
65	KELANTAN	INSTITUT BIMBINGAN AKADEMIK
66	KELANTAN	INSTITUT PENDIDIKAN
67	KELANTAN	INSTITUT SMAT
68	KELANTAN	INSTITUT TUNAS BAKTI
69	KELANTAN	INSTITUT ZAABA
70	KELANTAN	SEKOLAH TINGGI WADI SOFIA
71	JOHOR	MAKTAB ADABI MAHARANI
72	JOHOR	SEKOLAH MENENGAH ATAS (2) SEMAI
73	JOHOR	SEKOLAH MENENGAH SERI OMEGA
74	JOHOR	SEKOLAH MENENGAH SRI UTAMA
75	JOHOR	SEKOLAH MENENGAH SRI CAHAYA
76	JOHOR	SEKOLAH MENENGAH SRI ARA

77	KEDAH	SEKOLAH MENENGAH TENGKU MAHMOOD
78	KEDAH	SEKOLAH MENENGAH LANGKASUKA
79	MELAKA	KOLEJ YAYASAN SAAD
80	MELAKA	SEKOLAH METHODIST WESLEY MELAKA
81	N.SEMBILAN	KOLEJ TUANKU JAAFAR (SEK MEN)
82	N.SEMBILAN	SEKOLAH MENENGAH BERSEPADU KEMAYAN
83	N.SEMBILAN	SEKOLAH METHODIST WESLEY SEREMBAN
84	P.PINANG	SEKOLAH MENENGAH SRI TENBY
85	P.PINANG	SEKOLAH MENENGAH SRI PINANG
86	P.PINANG	SEKOLAH TINGGI SRI OLYMPIA
87	TERENGGANU	YAYASAN TERENGGANU CAW. KUALA BERANG
88	TERENGGANU	YAYASAN TERENGGANU CAW.BESUT
89	K.LUMPUR	SEKOLAH MENENGAH SRI DASMESH

90	K.LUMPUR	SEKOLAH MENENGAH SRI GARDEN
91	K.LUMPUR	SEKOLAH MENENGAH SRI SEMPURNA
92	K.LUMPUR	SEKOLAH MENENGAH STELLA MARIS
93	K.LUMPUR	SEKOLAH METHODIST WESLEY
94	K.LUMPUR	SEKOLAH SRI BUNGA RAYA
95	K.LUMPUR	SEKOLAH SRI CEMPAKA
96	K.LUMPUR	SEKOLAH SRI UTAMA KUALA LUMPUR
97	K.LUMPUR	SEKOLAH MENENGAH SRI FAIRVIEW
98	LABUAN	SEKOLAH MENENGAH SRI LABUAN
99	SABAH	MAKTAB NASIONAL
100	SABAH	SEKOLAH MENENGAH ADVENT GOSHEN
101	SABAH	SEKOLAH MENENGAH SWASTA SERI INSAN
102	SABAH	SEKOLAH MENENGAH ADVENT TAMPARULI

103	SABAH	SEKOLAH MENENGAH AL-SALAM
104	SABAH	SEKOLAH MENENGAH VISI
105	SARAWAK	SEKOLAH MENENGAH AYER MANIS
106	SARAWAK	SEKOLAH MENENGAH TUNKU PUTRA
107	SARAWAK	SEKOLAH MENENGAH LODGE
108	SARAWAK	SEKOLAH MENENGAH SUNNY HILL
109	PERAK	SEKOLAH RENDAH ISLAM PINTAR IPOH
110	SELANGOR	SEKOLAH RENDAH HAFIZ
111	SELANGOR	SEKOLAH RENDAH ISLAM PINTAR AL-AMIN
112	SELANGOR	SEKOLAH RENDAH SRI AL-AMIN BANGI
113	SELANGOR	SEKOLAH RENDAH IBNU KHALDUN
114	SELANGOR	SEKOLAH RENDAH ISLAM ADNI
115	SELANGOR	SEKOLAH RENDAH ISLAM HIRA’

116	SELANGOR	SEKOLAH RENDAH ISLAM AL HUDA
117	SELANGOR	SEKOLAH RENDAH ISLAM AL AMIN, GOMBAK
118	SELANGOR	SEKOLAH RENDAH ISLAM AL-AMIN
119	PAHANG	SEKOLAH RENDAH ISLAM INDERA MAHKOTA
120	PAHANG	SEKOLAH RENDAH ISLAM AR-IRSYAD
121	JOHOR	SEKOLAH RENDAH ISLAM AT-TAHFIZ
122	JOHOR	SEKOLAH RENDAH ISLAM HIDAYAH
123	JOHOR	SEKOLAH SRI JOHOR BAHRU
124	JOHOR	SEKOLAH RENDAH ISLAM HUSNI AMAL
125	KELANTAN	SEKOLAH RENDAH ISLAM AN-NUR
126	KELANTAN	SEKOLAH ISLAM AL-HIKMAH
127	KELANTAN	SEKOLAH RENDAH ISLAM AMAN
128	KELANTAN	SEKOLAH RENDAH ISLAM SRI KOTA

129	KELANTAN	SEKOLAH RENDAH ISLAM KIBLAH
130	KELANTAN	SEKOLAH RENDAH ISLAM AT- TAQWA
131	KELANTAN	SEKOLAH ISLAM (ARAB) DINIAH
132	KELANTAN	SEKOLAH RENDAH AGAMA (ARAB) AL-ITTIHADIAH
133	KELANTAN	SEKOLAH RENDAH AGAMA AL-QARI
134	KEDAH	SEKOLAH RENDAH AL-AZHAR
135	KEDAH	SEKOLAH RENDAH AL-ISLAH
136	KEDAH	SEKOLAH RENDAH AL-NAIM
137	KEDAH	SEKOLAH RENDAH ISLAM ABIM
138	KEDAH	SEKOLAH RENDAH ISLAM ALOR SETAR
139	KEDAH	SEKOLAH RENDAH ISLAM DARUL AMAN (SRIDA)
140	N.SEMBILAN	SEKOLAH RENDAH ISLAM (SRI) SEREMBAN
141	P.PINANG	SEKOLAH RENDAH AL-ITQAN

142	TERENGGANU	SEKOLAH RENDAH ISLAM AL-AMIN
143	PERAK	SEKOLAH MENENGAH IRSHADIAH
144	SELANGOR	SEKOLAH KIBLAH
145	SELANGOR	SEKOLAH MENENGAH ISLAM AL-AMIN GOMBAK
146	SELANGOR	SEKOLAH MENENGAH SRI AL-AMIN
147	SELANGOR	SEKOLAH MENENGAH IBNU KHALDUN
148	SELANGOR	SEKOLAH MENENGAH ISLAM HIRA'
149	SELANGOR	SEKOLAH MENENGAH ADNI
150	SELANGOR	SEKOLAH MENENGAH HAFIZ
151	JOHOR	SEKOLAH MENENGAH ISLAM AT-TAHFIZ
152	JOHOR	SEKOLAH MENENGAH ISLAM HIDAYAH
153	KELANTAN	SEKOLAH MENENGAH ISLAM AL-HUSNA
154	KELANTAN	SEKOLAH MENENGAH ISLAM AMAN

155	KEDAH	SEKOLAH MENENGAH AGAMA ISLAH
156	KEDAH	SEKOLAH MENENGAH AL-AZHAR
157	N.SEMBILAN	SEKOLAH MENENGAH SERI SETIA
158	N.SEMBILAN	SEKOLAH TINGGI ISLAM AS-SOFA
159	PERLIS	SEKOLAH MENENGAH AGAMA MA'HAD ATTARBIYAH AL-ISLAMIYAH (MATRI)
160	TERENGGANU	SEKOLAH MENENGAH ISLAM AL-AMIN KEMAMAN
161	PAHANG	MAAHAD TAHFIZ NEGERI PAHANG
162	TERENGGANU	SEKOLAH MENENGAH SRI UTAMA, K.TERENGGANU
163	SELANGOR	SEKOLAH RENDAH SERI SURIA
164	SELANGOR	SEKOLAH SRI TENBY(RENDAH)
165	SELANGOR	SEKOLAH MENENGAH SETIABUDI
166	N.SEMBILAN	KOLEJ TUANKU JAAFAR (SEK MEN)
167	KELANTAN	SEKOLAH ADABI PASIR PUTEH

168	JOHOR	SEKOLAH MENENGAH AUSTIN HEIGHTS JB
169	JOHOR	SEKOLAH RENDAH SERI OMEGA
170	JOHOR	SEKOLAH SRI JOHOR BAHRU
171	JOHOR	SEKOLAH RENDAH SRI ARA
172	JOHOR	SEKOLAH RENDAH SRI CAHAYA
173	JOHOR	SEKOLAH RENDAH AUSTIN HEIGHTS
174	JOHOR	PERSATUAN KANAK-KANAK SPASTIK JOHOR
175	P.PINANG	SEKOLAH SINAR HARAPAN
176	P.PINANG	SEKOLAH KANAK-KANAK TERENCAT AKAL P.PINANG
177	P.PINANG	PERSATUAN KANAK-KANAK SPASTIK P.PINANG
178	KELANTAN	PUSAT ASUHAN HARIAN KANAK-KANAK SPASTIK KELANTAN

English medium private schools in Malaysia-2011

BIL	NEGERI	NAMA SEKOLAH
1	PERAK	SEKOLAH ANTARABANGSA IPOH
2	SELANGOR	INTERNATIONAL ISLAMIC SCHOOL
3	SELANGOR	INTERNATIONAL SCHOOL OF KUALA LUMPUR (PRIMARY & SECONDARY)
4	SELANGOR	INTERNATIONAL SCHOOL OF KUALA LUMPUR (ELEMENTRY)
5	SELANGOR	AUSTRALIAN INTERNATIONAL SCHOOL MALAYSIA (PRIMARY & SECONDARY)
6	SELANGOR	MUTIARA INTERNATIONAL GRAMMAR SCHOOL
7	SELANGOR	SEKOLAH ALICE SMITH
8	SELANGOR	SEKOLAH ANTARABANGSA ELC
9	SELANGOR	MAZ INTERNATIONAL SCHOOL
10	SELANGOR	CEMPAKA INTERNATIONAL SCHOOL
11	SELANGOR	SUNWAY INTERNATIONAL SCHOOL (PRIMARY & SECONDARY)
12	SELANGOR	SRI KUALA LUMPUR INTERNATIONAL SCHOOL

13	SELANGOR	TANARATA INTERNATIONAL SCHOOL
14	SELANGOR	REAL INTERNATIONAL SCHOOL
15	SELANGOR	TENBY SCHOOL SETIA ECO PARK
16	SELANGOR	SRI KDU IB WORLD SCHOOL
17	SELANGOR	BRITISH INTERNATIONAL SCHOOL
18	SELANGOR	REAL INTERNATIONAL SCHOOL
19	PAHANG	INTERNATIONAL SCHOOL OF KUANTAN
20	PAHANG	GARDEN INTERNATIONAL SCHOOL
21	PAHANG	HIGHLANDS INTERNATIONAL BOARDING SCHOOL
22	KELANTAN	KELANTAN INTERNATIONAL SCHOOL
23	KELANTAN	WADI SOFIA INTERNATIONAL SCHOOL
24	JOHOR	UTAMA INTERNATIONAL SCHOOL
25	JOHOR	FAIRVIEW INTERNATIONAL SCHOOL

26	JOHOR	SERI OMEGA INTERNATIONAL SCHOOL
27	JOHOR	REAL INTERNATIONAL SCHOOL JOHOR
28	JOHOR	AUSTIN HEIGHT INTERNATIONAL SCHOOL
29	KEDAH	ELC INTERNATIONAL SCHOOL (KULIM)
30	MELAKA	MELAKA INTERNATIONAL SCHOOL
31	MELAKA	SEKOLAH EKSPATRIAT MELAKA
32	N.SEMBILAN	KOLEJ TUANKU JAAFAR
33	N.SEMBILAN	CEMPAKA INTERNATIONAL LADIES' COLLEGE
34	N.SEMBILAN	NILAI INTERNATIONAL COLLEGE
35	P.PINANG	ST CHRISTOPHER INTERNATIONAL PRIMARY SCHOOL, PENANG
36	P.PINANG	DALAT INTERNATIONAL SCHOOL
37	P.PINANG	INTERNATIONAL SCHOOL OF PENANG (UPLANDS)
38	P.PINANG	TENBY INTERNATIONAL SCHOOL

39	P.PINANG	FAIRVIEW INTERNATIONAL SCHOOL
40	TERENGGANU	EKHLASS INTERNATIONAL SCHOOL
41	TERENGGANU	SEKOLAH ANTARABANGSA UTAMA, TERENGGANU
42	PUTRAJAYA	NEXUS INTERNATIONAL SCHOOL
43	K.LUMPUR	SEKOLAH ANTARABANGSA FAIRVIEW
44	K.LUMPUR	SEKOLAH ANTARABANGSA GARDEN
45	K.LUMPUR	MON'T KIARA INTERNATIONAL SCHOOL
46	K.LUMPUR	SEKOLAH ANTARABANGSA SAYFOL
47	K.LUMPUR	UTAMA INTERNATIONAL SCHOOL K.LUMPUR
48	K.LUMPUR	THE ALICE SMITH SCHOOL
49	K.LUMPUR	GLOBAL INDIAN INTERNATIONAL SCHOOL
50	K.LUMPUR	VIKAS INTERNATIONAL SCHOOL
51	K.LUMPUR	THE INTERNATIONAL MODERN ARABIC SCHOOL

52	K.LUMPUR	SERI GARDEN INTERNATIONAL SCHOOL
53	LABUAN	SEKOLAH ANTARABANGSA LABUAN
54	SABAH	SEKOLAH ANTARABANGSA KINABALU
55	SABAH	SEKOLAH ANTARABANGSA SAYFOL
56	SARAWAK	LODGE INTERNATIONAL SCHOOL
57	SARAWAK	SEKOLAH ANTARABANGSA TUNKU PUTRA
58	SARAWAK	KIDURONG INTERNATIONAL SCHOOL
59	SELANGOR	ELC INTERNATIONAL SCHOOL (CYBERJAYA)
60	SELANGOR	FAIRVIEW INTERNATIONAL SCHOOL (SUBANG JAYA)
61	N.SEMBILAN	KOLEJ TUANKU JAAFAR

Appendix E

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Fax: +64 3 364 2069

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Information Sheet for parents (school students)

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

My name is Hasliza Abdul Halim. I am a PhD student at the Linguistics Department, University of Canterbury. I am conducting a research project that looks at ***restricted verb- phrase collocations in standard and learner Malaysian English.*** I am also interested to study the influence of exposure to English on the acquisition of Malaysian English restricted collocations. For examples, speakers of standard English usually say that students *sit an exam* not *undergo an exam*; or that someone *kept a diary*, rather than they *wrote a diary*. Your child's participation in the cloze test and his/her in-class written assignment will make an important contribution to this research.

I would like to invite your child to participate in my present study. If you agree to allow him/her to take part he/she will be asked to do the following:

- Fill in the blanks in a story and give their age, gender and whether or not they are a native speaker of English. Their name will not be known. This will take approximately 30-40 minutes.
- Write a short essay, approximately about 250-300 words. This will take 45-60 minutes.

Please note that participation in this study is voluntary. If your child participates, he/she has the right to withdraw from the study at any time without penalty. If he/she withdraws, I will do my best to remove any information relating to him/her, provided this is practically achievable.

I will take particular care to ensure the confidentiality of all data gathered for this study. I will also take care to ensure your child's confidentiality in publications of the findings. All the data will be securely stored in password protected facilities and locked storage at the University of Canterbury for ten years following the study. It will then be destroyed.

Please note that a PhD is a public document via the University of Canterbury Library database.

The results of the study may be submitted for publication to national or international journals or presented at educational conferences. You may at any time ask for additional information or results from the study.

If you would like more information or have any questions about the research, you can contact me or my supervisor Professor Koenraad Kuiper (kon.kuiper@canterbury.ac.nz). If you have any concerns or complaints about this research, please use the contact details below.

If you are happy to grant permission for your child to take part you will need to sign the consent form and return it to me in the envelope provided. Please retain this information sheet. Thank you for your consideration on this research project.

HASLIZA ABDUL HALIM (hasliza.abdhalim@pg.canterbury.ac.nz)
Office Phone: (03) 3667001 etxn: 8523.

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Consent Form for parents

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

I understand the aims and purposes of the research study undertaken by ***HASLIZA ABDUL HALIM.***

- ✓ The study has been explained to me and I understand the information that was given to me on the information sheet.
- ✓ I am aware that my child can withdraw from the study at any time without penalty, and he/she does not have to give any reason for withdrawing.
- ✓ I understand that his/her involvement will include completing a cloze test and producing a written assignment (essay).
- ✓ I understand that all information will be treated in strictest confidence, that participants will remain confidential and that no information that could identify my child will be given to other researchers or agencies. I understand that all data from this research will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for ten years following the study.
- ✓ I understand that a PhD is a public document via the University of Canterbury Library database.
- ✓ I understand that within these restrictions, the findings may be used to prepare articles for publication in national and/or international journals and for presentation at conferences; that the results of the study can be made available to me at my request and that I can request additional information at any time.
- ✓ I understand that I may receive either a copy of the full report or a summary of the findings of this study and have provided my email details below for this purpose. I realise that whether or not I decide to allow my child to participate is my decision and will not affect his/her grade.
- ✓ I agree to grant permission for my child to participate in this research.

Name :_____

Child's name :_____

Class :_____

Signed :_____

Date :_____

Email address for report: _____

[Note: The parents/caregivers will also receive a full information sheet and will be required to complete a consent form as well before the child can take part in this research.]

Please ***return this completed consent form*** *in the envelope provided.*

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Fax: +64 3 364 2069

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Information Sheet for the participants (school students)

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

My name is Hasliza Abdul Halim. I am a PhD student at the Linguistics Department, University of Canterbury. I am conducting a research project that looks at ***restricted verb- phrase collocations in standard and learner Malaysian English.*** I am also interested to study the influence of exposure to English on the acquisition of Malaysian English restricted collocations. For examples, speakers of standard English usually say that students *sit an exam* not *undergo an exam*; or that someone *kept a diary*, rather than they *wrote a diary*. Your participation in the cloze test and your in-class written assignment will make an important contribution to this research.

I would like to invite you to participate in my present study. If you agree to take part you will be asked to do the following:

- Fill in the blanks in a story and give your age, gender and whether or not you are a native speaker of English. Your name will not be known. This will take approximately 30-40 minutes.
- Write a short essay, approximately about 250-300 words. This will take 45-60 minutes.

Please note that participation in this study is voluntary. If you do participate, you have the right to withdraw from the study at any time without penalty. If you withdraw, I will

do my best to remove any information relating to you, provided this is practically achievable.

I will take particular care to ensure the confidentiality of all data gathered for this study. I will also take care to ensure your confidentiality in publications of the findings. All the data will be securely stored in password protected facilities and locked storage at the University of Canterbury for ten years following the study. It will then be destroyed.

Please note that a PhD is a public document via the University of Canterbury Library database.

The results of the study may be submitted for publication to national or international journals or presented at educational conferences. You may at any time ask for additional information or results from the study.

If you would like more information or have any questions about the research, you can contact me or my supervisor Professor Koenraad Kuiper (kon.kuiper@canterbury.ac.nz). If you have any concerns or complaints about this research, please use the contact details below.

If you are happy to take part you will need to sign the consent form and return it to me in the envelope provided. Please retain this information sheet. I am looking forward to working with you and thank you in advance for your contributions.

HASLIZA ABDUL HALIM (hasliza.abdhalim@pg.canterbury.ac.nz)
Office Phone: (03) 3667001 etxn: 8523.

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Fax: +64 3 364 2069

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Consent Form for school students

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

I understand the aims and purposes of the research study undertaken by ***HASLIZA ABDUL HALIM.***

- ✓ The study has been explained to me and I understand the information that was given to me on the information sheet.
- ✓ I am aware that I can withdraw from the study at any time without penalty, and I do not have to give any reason for withdrawing. I have had all questions answered to my satisfaction.
- ✓ I understand that my involvement will include completing a cloze test and producing a written assignment (essay).
- ✓ I understand that all information will be treated in strictest confidence, that participants will remain confidential and that no information that could identify me will be given to other researchers or agencies. I understand that all data from this research will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for ten years following the study.
- ✓ I understand that a PhD is a public document via the University of Canterbury Library database.
- ✓ I understand that within these restrictions, the findings may be used to prepare articles for publication in national and/or international journals and for presentation at conferences; that the results of the study can be made available to me at my request and that I can request additional information at any time.
- ✓ I understand that I may receive either a copy of the full report or a summary of the findings of this study and have provided my email details below for this purpose. I realise that whether or not I decide to participate is my decision and will not affect my grade.
- ✓ I agree to participate in this research and my parents have also given consent on their consent form.

Name :_____

Class :_____

Class teacher :_____

Signed :_____

Date :_____

Email address for report: _____

[Note: The parents/caregivers will also receive a full information sheet and will be required to complete a consent form as well before the child can take part in this research.]

Please ***return this completed consent form*** *in the envelope provided.*

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Fax: +64 3 364 2069

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Information Sheet for the participants (university students)

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

My name is Hasliza Abdul Halim. I am a PhD student at the Linguistics Department, University of Canterbury. I am conducting a research project that looks at ***restricted verb- phrase collocations in standard and learner Malaysian English.*** I am also interested to study the influence of exposure to English on the acquisition of Malaysian English restricted collocations. For examples, speakers of standard English usually say that students *sit an exam* not *undergo an exam*; or that someone *kept a diary*, rather than they *wrote a diary*. Your participation in the cloze test and your in-class written assignment will make an important contribution to this research.

I would like to invite you to participate in my present study. If you agree to take part you will be asked to do the following:

- Fill in the blanks in a story and give your age, gender and whether or not you are a native speaker of English. Your name will not be known. This will take approximately 30-40 minutes.
- Write a short essay, approximately about 250-300 words. This will take 45-60 minutes.

Please note that participation in this study is voluntary. If you do participate, you have the right to withdraw from the study at any time without penalty. If you withdraw, I will

do my best to remove any information relating to you, provided this is practically achievable.

I will take particular care to ensure the confidentiality of all data gathered for this study. I will also take care to ensure your confidentiality in publications of the findings. All the data will be securely stored in password protected facilities and locked storage at the University of Canterbury for ten years following the study. It will then be destroyed.

Please note that a PhD is a public document via the University of Canterbury Library database.

The results of the study may be submitted for publication to national or international journals or presented at educational conferences. You may at any time ask for additional information or results from the study.

If you would like more information or have any questions about the research, you can contact me or my supervisor Professor Koenraad Kuiper (kon.kuiper@canterbury.ac.nz). If you have any concerns or complaints about this research, please use the contact details below.

If you are happy to take part you will need to sign the consent form and return it to me in the envelope provided. Please retain this information sheet. I am looking forward to working with you and thank you in advance for your contributions.

HASLIZA ABDUL HALIM (hasliza.abdhalim@pg.canterbury.ac.nz)
Office Phone: (03) 3667001 etxn: 8523.

Department of Linguistics

School of Languages, Cultures and Linguistics

University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Fax: +64 3 364 2069

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Consent Form for university students

Project title: **Restricted Verb-Phrase Collocations in Standard and Learner Malaysian English**

I understand the aims and purposes of the research study undertaken by ***HASLIZA ABDUL HALIM.***

- The study has been explained to me and I understand the information that was given to me on the information sheet.
- I am aware that I can withdraw from the study at any time without penalty, and I do not have to give any reason for withdrawing. I have had all questions answered to my satisfaction.
- I understand that my involvement will include completing a cloze test and producing a written assignment (essay).
- I understand that all information will be treated in strictest confidence, that participants will remain confidential and that no information that could identify me will be given to other researchers or agencies. I understand that all data from this research will be securely stored in password protected facilities and/or locked storage at the University of Canterbury for ten years following the study.
- I understand that a PhD is a public document via the University of Canterbury Library database.
- I understand that within these restrictions, the findings may be used to prepare articles for publication in national and/or international journals and for presentation at conferences; that the results of the study can be made available to me at my request and that I can request additional information at any time.
- I understand that I may receive either a copy of the full report or a summary of the findings of this study and have provided my email details below for this purpose. I realise that whether or not I decide to participate is my decision and will not affect my grade.
- By signing below, I agree to participate in this research project.

Name :_____

Signed :_____

Date :_____

Email address for report: _____

*Please **return this completed consent form** in the envelope provided.*

Appendix F

Thank you very much for agreeing to take part in this survey conducted by PhD student Hasliza Abdul Halim from the School of Languages, Cultures and Linguistics of the University of Canterbury, New Zealand.

It is important for you to know that this is not a test but a survey. The survey is about your knowledge of English and how that is influenced by various things. There are no right or wrong answers, and you do not even have to write your name on it.

The results of this survey will be used only for research purposes so please answer honestly. Please read each item carefully before responding. The information we get from you will be vital to the success of this project.

1. Age : _____ (Please do not write your name)

2. Gender :_____

3. What is your native language? English ( ) Malay ( )
Mandarin ( ) Tamil ( )
Other ( ) Please specify:_____

4. If English is not your native language, how long you have been learning/ speaking English?

_____years, _____months

5.What is your score for a recent test of IELTS/ TOEFL/ MUET/other examinations:

_____

6.What was your recent or last school attended?

_____

7.What language did your teachers use when they were teaching subjects other than English at

your current or previous school?_____

8.What is the main language you use every day for these occasions?

a) formal (in class at school)-_____

b) informal/casual (reading, facebook, TV programmes, conversations)-_____

This survey comprises only **ONE** section: **FILLING IN THE GAPS**

Please make sure you complete the section otherwise the survey will not have complete results. So just put in your best guess in the gap even if you are not sure of what word you think ought to go in the gap.

Thank you for taking the time to help us with our research.

Please read the story below and when you find a gap (____), write in the verb you think should go there.

For example:

Tomorrow the students will <u>sit</u> the exam.

Don't worry if you can't think of a word straight away, just put in your best guess.

AN AUSPICIOUS NIGHT ESCAPE AT PETALING STREET

If you are planning to celebrate Chinese New Year in Kuala Lumpur you should know that you could not have a better place than Chinatown of Kuala Lumpur, also known as Petaling Street. This part of the city is prosperously decked in Chinese New Year decorations and the vibrant colour of the festive mood. These are all set up to welcome any visitors from any part of the world. Petaling Street is the traditional hub for Chinese New Year countdown every year. Many Malaysian Chinese come here to buy food, flowers and decorative items for the festive season. Among all visitors you can spot many teenagers among the bustling groups who come here for window shopping. I was among these buzzing groups of people who like to entertain ourselves looking at interesting things sold in the stalls. It was also interesting looking at visitors did some shopping and bargaining for cheaper deals. The visit to Petaling Street was a 'must do' activity for at least once in a fortnight. That was about ten years ago when I **(1)** _____**the responsibility** of being a local university student in Kuala Lumpur. And this year I decided to visit one of my favourite places in Kuala Lumpur.

With great excitement, my two friends Pei Lie and Kavitha suggested, "Let's go to Chinatown!" echoing my enthusiasm to 're-experience' the festive mood in KL Chinatown.

Kavitha and I decided to pick Pei Lie up from her house at 4p.m. We stopped in front of a green painted double storey house. We spotted Uncle Chan was in the garden smiling at us. We nodded and waved at him. We saw Pei Lie rushing out and waved at us.

"Do you think it's going to rain, girls? I think I should bring umbrella with me." said Pei Lie.

"It's fine Pei Lie. I do have extra umbrellas in the car. Come on, we are running late," Kavitha said.

While we were waiting for Pei Lie we heard Pei Lie's 12 year old sister, Mei Mei yelled from the front window, "Don't forget my Kung Fu chestnuts. I want two bags. You better buy them for me if not I will not record your favourite movie tonight."

"I don't **(2)** _____**a hoot** whether you're willing to record that or not because I've already told Mum to do that for me," Pei Lie shouted back.

"I don't care, I still want my Kung Fu chestnuts," insisted Mei Mei.

"Girls, **(3)** _____ **bickering** please," Aunty Nancy tried to calm them down. "I don't want neighbours to hear you. It would be so embarrassing if Aunty Noriah and Aunty Devi heard you girls fighting. Pei Lie, don't come back late. And don't forget my 'longan drink'. Kavitha and Lisa please say 'hi' to your mothers."

We waved at Aunty Nancy who was standing at the front door. We passed by Uncle Chan. We nodded, and said a quiet 'goodbye' and smiled to **(4)** _____**our respect**.

Kavitha was driving the car quietly. In fact we were all silent so I said something to break the ice, "Pei Lie, how's your Chinese New Year preparations? Have you bought new clothes and finished decorating your house. I believe you and your family are pretty busy right now."

"Hmm.."Pei Lie sighed. "We are partially done. As you know in Chinese tradition the preparation should begin a month prior to the New Year. We have to clean the house from top to bottom, then we decorate the house with red lanterns, banners and paper or plastic fire crackers. We also have to put up panels inscribed with calligraphic characters bearing theme of happiness, wealth and longevity."

"It's only 2 days left to the New Year. I believe your mother must be the busiest person on earth doing all the preparations," added Kavitha smiling.

"Oh, she is. You know what she's very rigid in terms of traditions and norms of our cultures and beliefs. As you know there are **taboos and prohibited things** that we should **(5)** _____, otherwise misfortune may befall the family. A funny yet still strictly followed example is no one is allowed to sweep the floor on the first day of the New Year as it is considered unlucky. My mother always reminds us that, "one would accidently sweep away one's good luck and fortune if they do so," explained Pei Lie.

"Wow, your mother is a faithful believer and the great fact is that she takes a firm stand to ensure your traditions are respected, "I added.

"Yeah, there are fascinating aspects of Chinese New Year tradition and rituals that I admire. What the ancestors started in the past begin and still continue to take root in the younger generation like Mei Mei," said Kavitha while her eyes focusing on the busy road. "I don't like the traffic jam in KL. It really gives me headache."

"Hei, what about the 'makan besar' or reunion dinner" I asked Pei Lie curiously. "Is that dinner takes place on the eve of New Year?"

"The eve of New Year is considered the high point of the celebration as it is on that particular day that all family members from far or near will return home for the reunion dinner, to **(6)** _____ **family ties** and to revive family relationship. Normally our dinner is made up of seafood and dumplings. We also serve ducks, prawns, braised dried oysters, scallops and 'yeesang' or prosperity vegetables. And after dinner, family members will try to stay up all night in adherence to 'shou sui', a practice which is believed to bring one's parents longevity. So, to **(7)** _____ **time**, some members would play mahjongs. And at the stroke of midnight, you'll hear the firecrackers and the fireworks explode," explained Pei Lie enthusiastically.

"Fireworks and firecrackers are the best part of all," I jumped excitedly.

"Yes, yes, me too," Kavitha agreed.

Suddenly,"Oh no!" Kavitha shouted and suddenly brake the car. We heard a loud bang in front of us. And suddenly people started to crowd the place.

As Kavitha was shivering, Pei Lie tried to calm her, "Stay calm Kavitha, there's no need to press the panic button. We are all safe. That's good enough."

"I could see a four wheeler, a Honda I guess hit a motorcyclist," I explained. I could see Kavitha was **(8)** _____ **her teeth** in frustration over what had happened just now. "It's ok, Kavitha. I suggest we should take left turn to Petaling Street if not we are going to stuck here forever." I felt the need to **(9)** _____ **my view** as well.

"Yeah, this is KL.' I said, vent my frustration at the delay. "You can easily be hit even by a bus. Ahh, the traffic is so frustrating," sighed Pei Lie. "Don't worry, girls. We're just about 5 minutes to our destination." She tried to cheer us up.

We parked our car near Puduraya. It was only 2-3 minutes walking to Petaling Street. We were just three of the many visitors who went there at this time of the year just for the feel of the festive excitement. Originally, Petaling Street was known as 'Chee Cheong Gai' in Cantonese, or 'Starch Factory Street'. It was the settlement of Chinese workers from China since 140 years ago.

"Wow, it looks very much different," I murmured to myself.

"Lisa, don't you know that in 2003 this area has been refurbished, upgraded and some kind of 'reimaged' as Chinatown of KL. But, the project was left unfinished for few months. Thank God, the City Council stood up and quickly **(10)** _____ **the backlog** of work on the project and finished it. I'm so happy the City Council took a very drastic action," explained Kavitha.

"Lisa, you could **(11)** _____ **the difference** between the way the street was and now just from the two new large Chinese arches. They are constructed at both end of the street to welcome visitors, and the good thing is that the flea market has been properly organised and rearranged for a more tidy looking stalls," added Pei Lie.

"It is good as the hawkers really benefitted from the renovations. They're making more money as there are more tourists visiting every day. It is really give the street a big boost," explained Kavitha excitedly.

"Ha, ha, ha. Don't forget those who are **(12)** _____ **a fast buck** selling prohibited items like firecrackers and fake branded items like bags, watches and other things," said Pei Lie. She further added, "My biggest interest in Petaling Street are its local meat and snacks and not to forget the assam laksa. Aaah, so nyum, nyum," she said while rubbing her flat tummy. Kavitha and I could not stop laughing at her action.

We could see hawkers have decorated their stalls with red lanterns, some red and gold banners and firecrackers made of plastic or papers. The night looked so amazingly bright. We passed by a pomelo stall. Pei Lie bought some home.

We took our time to walk along the paved foot way. We were really enjoying the sights, sounds and colours of prosperity. Then, we came to our favourite stall of all, 'longan' or 'dragon-eye' drink stall. This is the most popular stall in Petaling Street. The good thing is that the stall offered a very good deal of 1.20 Ringgit Malaysia per cup. It was so irresistible. I believe every visitor deserves a try. I had to **(13)** _____ **a grip of myself** for not getting a second cup of that drink although people say it **(14)** _____ **wonders** for your complexion. As many people were crowding the stall, Pei Lie still got the chance to have a chat with the stall owner who was her family friend. She introduced us to him, and we exchanged smiles and stories. Suddenly, a loud drumming sound tried to **(15)** _____ **the show**. We were taken aback by a group performance of the lion dancers. The lion with its fierce features certainly **(16)** _____ **the part**. It was a vigorous form of dance that usually performs during Chinese New Year festivals.

Pei Lie whispered, "I guess they're practicing for the festive day performance. We're lucky today, girls".

The longan drink stall owner tried to **(17)** _____ **a joke**, "Be careful girls, the lion might eat you up. All of us started to laugh.

"Oh, don't worry girls. The lion will not eat us all if Kavitha **(18)** _____ **the bill** for the drinks," I said. We burst out laughing again.

That was not the first time I witness the traditional dance in front of my very own eyes. The only difference was that this time I was standing quite close to the dancers. Yes, I guess I was pretty lucky. I was delighted to see how the dancers danced in full co-ordinations and elegance. I was so amazed of how this dance lit up the auspicious day of Chinese New Year. It was quite exciting looking at how the dancers gave 'life' to a lion where one of them had to control the movements of the head, eyes and mouth, and the rest to act and move around as the body.

"It is so amazing! I know it is not an easy task to lift the lion's head which is adorned with feathers, furs and glitters. I believe it weighs from 8-12kgs," Kavitha half shouted as the music and drumming were to loud.

"You know what, at the end of the show the dancers are usually enticed with gifts. They would normally be 'angpows' (money in red packets) attached to a vegetable which a tied to a pole," explained Pei Lie. "I think we should move on and visit other stalls. I need to get Kung Fu chestnuts for my sister. I'm **(19)** _____ **a big risk** if forgetting my promise" uttered Pei Lie.

Kavitha replied, "Yes, better for you to **(20)** _____ **win-win situation** as she's helping you with the movie recording and you reward her with Kung Fu chestnuts in return. What a good deal. It's pretty fair, right?"

I smiled. I know Pei Lie is a good sister and daughter. She's been helping her father to put food on table by working hard at her job. She has been my friend since five years ago. I adore her for her sincere love and great friendship.

"Ha, ha red is everywhere. It means happiness, right? Yes, that's what I really experience tonight," Kavitha half shouted. "Can we stop and get those as well?" pointing her finger at the hanging red lanterns.

We came to a stall selling traditional Chinese candies made from ginger, melon, plums and lotus seeds. We bought few packs home. Finally, we reached the Kung Fu chestnuts stall.

"This is a 'must eat food' in Petaling Street. I guess I should get myself more than a packs right, girls?" I could not wait but told the stall owner my order.

"Yes Lisa, go for it. It's your lucky day. I'm getting myself some as well," replied Pei Lie while taking out some money from her purse. Kavitha laughed and we bought home full handed of Kung Fu chestnuts.

We cruised home as the clock started to strike nine. We felt tired yet so delighted to rediscover our so called as ‘secret rendezvous’. Both our hands were full of auspicious goodies and we left the lane for its nostalgic feeling.

Thank you.

Answers.

1. **Shouldered** responsibility
2. **Give** a hoot
3. **Stop** bickering
4. **Pay** respect
5. **Observed** taboo/prohibited things
6. **Rekindle** family ties
7. **Kill** time
8. **Gnash** teeth
9. **Air** view
10. **Cleared** backlog
11. **Tell** the difference
12. **Making** a fast buck
13. **Get** a grip of myself
14. **Does** wonders
15. **Steal** the show
16. **Look** the part
17. **Crack** a joke
18. **Foot** the bill
19. **Taking** a big risk
20. **Create** a win-win situation

Appendix G

Thank you very much for agreeing to take part in this survey conducted by PhD student Hasliza Abdul Halim from the School of Languages, Cultures and Linguistics of the University of Canterbury, New Zealand.

This is not a test so there are no right or wrong answers, and you do not even have to write your name on it. The results of this survey will be used only for research purposes so please answer honestly. Please read each item carefully before responding. The information we get from you will be vital to the success of this project.

- ***Age*** *: _____* *(Please do not write your name)*
- ***Gender*** *:_____*

- ***What is your native language?***

  English ( ) Malay ( ) Mandarin ( )

  Tamil ( ) Other ( ) Please specify:_____

- ***If English is not your native language, how long you have been learning/ speaking English?***

  _____years, _____months

- ***Please choose the medium of instruction at the school (high school) you attended***

English *Malay*

This survey comprises only ONE section: CLOZE TEST

Please make sure you complete the section.

Thank you very much.

Thank you for taking the time to help us with our research.

Please read the story and when you find a gap (_____), write in the verb you think should go there.

For example:

Tomorrow the students will <u>sit</u> the exam.

Don't worry if you can't think of a word straight away, just put in your best guess.

Shannon walked into the vast, badly-decorated function room and looked around for faces she knew. Tom's hand waved frantically at her from a table near the coat check. *Thank goodness!* she thought to herself, *I'm not the only one from Data Entry here!* Shannon normally **_____ these kind of events like the plague**, but her pushy new flatmate had convinced her a night out might be in order, especially considering she'd only been at the firm for a few weeks. She walked towards Tom and was further relieved to see familiar faces from Accounts, which had its offices on the same floor as her department. On closer inspection, she realised she knew a few others at the table – Jenny, who had a tendency to **_____ the goat** at inter-departmental health and safety meetings, pulling faces and telling stupid jokes; Annabel, who always looks like a startled deer when you ask her anything that isn't work-related, her face going blotchy at the prospect of real conversation; and Jonno, who **_____ every woman on the floor the creeps** with his fake smile and lame innuendo. It was clear that he'd already found his target for the evening, singling out a youngish redhead opposite Jenny. *This could be more tedious than I expected* Shannon grumbled as she got to the table and took the empty seat next to Jonno. She said a quiet hello and waited for Tom to do the proper introductions.

"Shannon, you remember Jenny, Annabel and Jonno, don't you? That's Annabel's friend Kim over there, and this is Peter from Marketing, Peter, this is Shannon from Data Entry" Tom duly offered.

"Hello Shannon" Peter smiled warmly and shook her hand. *Not bad,* she thought, *the night might not be a waste of makeup after all.* She looked up just in time to hear Tom finish the next introduction.

"…ndy, from, erm, sorry, which division is it you're in again?"

"Stores", he said tersely, "and it's Andrew".

"Ah, right, sorry about that, Andrew. And, ah, this is Shannon, from Data Entry" Tom rushed. Shannon nodded, smiled wanly and mumbled something polite. *Blimey! If that's his friendly party manner then he'll* ***_____ the lot of us to drink****! Mind you, that's likely to happen, it being the Christmas do and all.* Shannon directed her attention back to Peter. He was looking at her too and trying to be heard over the deep and meaningful Jonno was having with the redhead. Shannon tried to make his words out. *Pray? Or something about a braid? A ray?*

"Shall we **_____ the fray**?" he half-shouted.

"Sorry?!!"

"Dance? You want to dance?" He pointed at the large group of people wriggling to the strains of the Macarena.

"Sure" she yelled back, happy to get the chance to chat.

On the dance floor, Peter started talking about their colleagues moving around them, pointing out various departmental managers and PAs.

"She's really into one of those party-plan companies," he started, nodding towards a brunette spinning around to Kylie Minogue, "so try not to get into a long chat with her or you'll end up with endless cooking products! She's fanatical about _____ **the company line**, though. And he's the assistant manager out at the warehouse, always trying to _____ **the others into action**. You know, the old 'Go team' speech." Peter sighed quietly. "Shame though, seeing as the 'team' don't exactly _____ **the ground he walks on**."

"Yeah, office politics must be the same the world over!" Shannon cringed inwardly as she heard the dross coming out of her mouth. *Typical, as soon as anyone decent starts talking to me I lose about 30 IQ points!*

"So how long have you been working here?" she ventured, hoping to make up for her dullness.

"Oh, ah…in the company a few years, but here in Manchester only the last eight months. And you?"

"Only a few weeks. I needed a change from call centre work". Peter grinned knowingly at that.

"Was that here in Manchester?"

"No, I'd been living in Australia for a while. I moved back and started here the next week"

"You don't _____ **things by halves**, do you!?", Peter laughed, "change jobs and change countries for a break!"

"I'll _____**you into a secret**" she smirked. "I wasn't exactly mad on staying in Australia. I just waited 'til after I'd had more than enough of my job to make the decision."

"Not your cup of tea, then? All that sun and sea and sand?" Peter looked a little puzzled. Shannon could tell he thought all of Australia must be a paradise on earth.

"There was sand, all right. Sand and more sand and sand and snakes and spiders and more sand again! I was in the 'back of beyond', as the Aussies would say. I guess I'm

not really a country girl. So I handed in my notice and _____ **tracks** for the safety of inner city England".

"Whatever makes you happy, I guess, though I don't get how you could leave those temperatures! Still, I don't think I'd have ever _____ **up the courage** to move over there in the first place. Manchester's about the most exotic place I've ever lived."

"Yeah, I think the snake in my bed after a bad day at work _____ **my fate** -- I booked my flight home the next day!"

"Yes, well, can't say I blame you for that!"

The music changed to something slower and they instantly moved apart, the uncomfortable moment echoed by other dance partners around the room.

"Shall we go back to the table for a bit?" Shannon nodded readily, following him back to the table. She knew she was beginning to _____ **a fancy to the guy**, and a little group conversation might help her from going overboard.

"Ah, there you are, Shannon! Looks like you two have been getting friendly. Nice dance, was it?" Her supervisor's comment turned her face bright red, and when she looked up Tom was desperately trying to _____ **a straight face** over the tactlessness of their boss. Peter, fortunately, seemed not to have heard.

"They've _____ **no expense**, have they?" The supervisor picked up the '99p shop' decorations in the centrepiece. "At least the food smells promising. Sorry, I'm Jake Lewis, 2IC in Data Entry. And you're?"

"Peter Mayell. Marketing manager." *Manager?! No wonder he had all the gossip!* "My team planned the party."

Tom caught Shannon's eyes, both enjoying the brief moment of discomfort Jake went through before Peter admitted he was pulling his leg. Everybody at the table had a good laugh, but Shannon knew Jake really wanted to _____ **Peter's neck**.

"Seriously, though, aren't they trying to _____ **their belts**? You know, less spending on frivolous parties and more on real staff benefits." Tom asked. Everyone groaned at the shoptalk and the group quickly found some meaningless politician's embarrassment to steer the direction away from work. They'd barely started on the relentlessness of the tabloids when the PA system came on to announce the CEO's presence and inevitable speech.

"Good evening everyone, and thank you for coming here tonight to celebrate another successful year at Smith Industries" he started. The dull hum around the room was proof of the lack of interest in the expected description of income, outgoings and profit

margins. A quick scan of the rest of her table told Shannon most of the others were equally indifferent, though Peter was missing. She cast her eyes over the neighbouring tables but couldn't spot him. Tom, meanwhile, was listening intently. Reluctantly, she focused her attention back on the speech.

"This year we're doing something a bit different. You may have noticed the slightly less-than-fancy decorations on the table and the jukebox in place of a DJ or band. If you were thinking that the evening looked a little cheap, well, it is. I mean, the party you can *see* is cheap. This year, however, we thought we'd **_____ last year's Christmas party off the map**! If you'd like to stand up and follow Mr Mayell and Mr Thomsen out into the hallway, you may find yourselves pleasantly surprised."

Amongst the murmurs of confusion, Shannon stood up to follow Peter out. *No wonder he'd joked about the party – he really had organised it!*

Everyone followed Peter's lead, into the hallway and then out onto the bus lay by outside. There were a few moans about the company **_____ the bottom of the barrel** with a bus trip before two airline coaches drew up and opened their doors.

Inside, Peter explained to the group that the company's heads felt the profits this year deserved a real 'thank you' to the staff. Instead of the usual buffet, they were in fact taking a champagne service charter plane to Edinburgh, or more rightly a boutique hotel near Edinburgh, and would be meeting all the full-timers from the other two company offices there for the company's first-ever 'proper' Christmas party. Shocked silence turned into cheers and Peter headed over to sit near Shannon.

"Jesse Thomsen's doing the same in the other coach," he explained, "and I couldn't tell anyone what we were doing, so sorry about telling your supervisor we were pulling his leg! We were, just in the 'this isn't really a party' kind of way." He laughed again and his eyes sparkled. "It's going to be a big night!"

She nodded in agreement. All Shannon could think about was how she was going to thank her flatmate for making her come out tonight. Peter leaned towards her and quietly murmured something about a dinner date the next weekend. The coach drew nearer the airport and the excited noises got louder. *Definitely not a waste of make up!* She smiled to herself. *I think I might get to quite like this job!*

The End.Thank you.

Answers

1. avoided NP of events like the plague
2. act the goat
3. gave NP on the floor the creeps
4. drive NP to drink
5. enter/join the fray
6. toe the company line
7. galvanise/goad/spur NP into action
8. worship the ground he walks on
9. do things by halves
10. let you into a secret
11. made tracks
12. pluck/summon up courage
13. sealed NP's fate
14. take a fancy to NP
15. keep a straight face
16. spared no expense
17. wring NP's neck
18. tighten NP's belts
19. wipe NP off the map
20. scrape the bottom of the barrel